AD-A247 632

92-06969

Persistence in Climate

G. MacDonald

A-\

February 1992

JSR-91-340

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE February 1992	3. REPORT TYPE AND DATES COVERED Technical

4. TITLE AND SUBTITLE

Persistence in Climate

5. FUNDING NUMBERS

PR - 8503Z

6. AUTHOR(S)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The MITRE Corporation
JASON Program Office, A020
7525 Colshire Drive
McLean, Virginia 22102-3481

8. PERFORMING ORGANIZATION REPORT NUMBER

JSR-91-340

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Department of Energy
Office of Energy Research, ER-30
Washington, DC 20585

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

JSR-91-340

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution unlimited; open for public release.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Persistence in weather forecasting is used to describe runs of several days with similar weather characteristics. This general notion of persistence is extended to long term records of climate by examining the scaling properties of the range, maximum minus minimum, of the integral or sum of observed or calculated variable. The values of persistence, P, are limited by existence considerations to $-1 \leq P \leq 1$. For $P = 0$, the increments making up the sum are uncorrelated, independent variables. Values of P near unity represent a tendency for long runs of similar values. Observed global average annual temperature records exhibit strong positive persistence even when linear trend is removed. A hundred year CCM run (CCM-1) shows vanishing persistence perhaps indicating that the real oceans give rise to runs of several years or decades with similar climate characteristics while the model ocean fixed by seasonal means does not.

14. SUBJECT TERMS

CCM, persistence, GCM

15. NUMBER OF PAGES

70

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

Abstract

Persistence in weather forecasting is used to describe runs of several days with similar weather characteristics. This general notion of persistence is extended to long term records of climate by examining the scaling properties of the range, maximum minus minimum, of the integral or sum of observed or calculated variable. The values of persistence, P, are limited by existence considerations to $-1 \leq P \leq 1$. For $P = 0$, the increments making up the sum are uncorrelated, independent variables. Values of P near unity represent a tendency for long runs of similar values. Observed global average annual temperature records exhibit strong positive persistence even when linear trend is removed. A hundred year CCM run (CCM-1) shows vanishing persistence perhaps indicating that the real oceans give rise to runs of several years or decades with similar climate characteristics while the model ocean fixed by seasonal means does not.

Contents

1 INTRODUCTION

Questions about predictability of climate focus attention on the phenomenon of persistence. In meteorology, the term "persistence" is used to describe runs of several days with similar weather characteristics. Persistence can aid weather forecasting: Under some circumstances today's weather can be used to predict tomorrow's. In the case of high frequency weather, persistence is associated with the large scale organized motion within the atmosphere, the nature of which is determined in large measure by the earth's rotation. For example, the air mass comprising a high pressure system, with anticyclonic circulation, has similar characteristics over the area it covers and weather conditions can persist for days as the system moves over a locality. Such a system may take several days to pass over a given location in summer, while in winter it moves more rapidly. Therefore, in general, predictive skill is greater in summer than in winter.

Does climate show persistence? Are there runs lasting several years or more in which similar conditions persist that are different from "mean" conditions? Does climate exhibit long term memory so that long term correlations in patterns of climate exist? Are long runs of drought recorded in Biblical times and in today's California examples of persistence in climate?

If climate does demonstrate these long term correlations in the form of runs, then this persistence may assist long term climate prediction. The physical origin for such persistence might be in the longer term organized motion of the ocean, where time scales of a thousand years may be expected, or in the complexities of nonlinear dynamics of both oceans and atmospheres.

1

While persistence is a familiar term in meteorology there is no quantitative definition nor any clear cut way of determining the degree of persistence in an observed or calculated time series. Long runs of similar values of a parameter describing climate, such as global annual average temperature, would reveal themselves as energy in the low frequency bands of the power spectrum. In fact, long meteorological (and other geophysical) records typically show abundant energy at low frequencies, often described as red noise. This behavior is illustrated in the cumulative spectrum (see Figure 1) of the average global temperature calculated from the Lorenz 27-variable model integrated over 26,000 years (see Section 4). About 90 percent of the variance lies in the lower 20 percent of the frequency domain. Century long observational temperature records show similar spectra (see Section 5). A current GCM with ocean temperatures fixed by seasonal means does not display observed low frequency behavior nor persistence (see Sections 5.2 and 5.3).

Classical time series statistics are weak in describing low frequency behavior, although the two point correlation function is useful in describing short time correlations. Higher order correlations are required to describe long term runs, but these quantities are very computationally intensive and do not lend themselves to easy graphical presentation or interpretation. Power spectra are notorious for their weakness in representing behavior near zero frequency. If the mean is removed, the zero value at zero frequency pulls down estimates of neighboring frequencies due to the "windowing" associated with the finite length of the record. If the mean is included, the delta function at zero frequency pulls up nearby estimates. Any linear trend is usually removed prior to taking the spectrum because the low frequency content of a trend distorts the estimates of the low frequency part of the spectrum. Since the number of "cycles" at periods comparable to the record
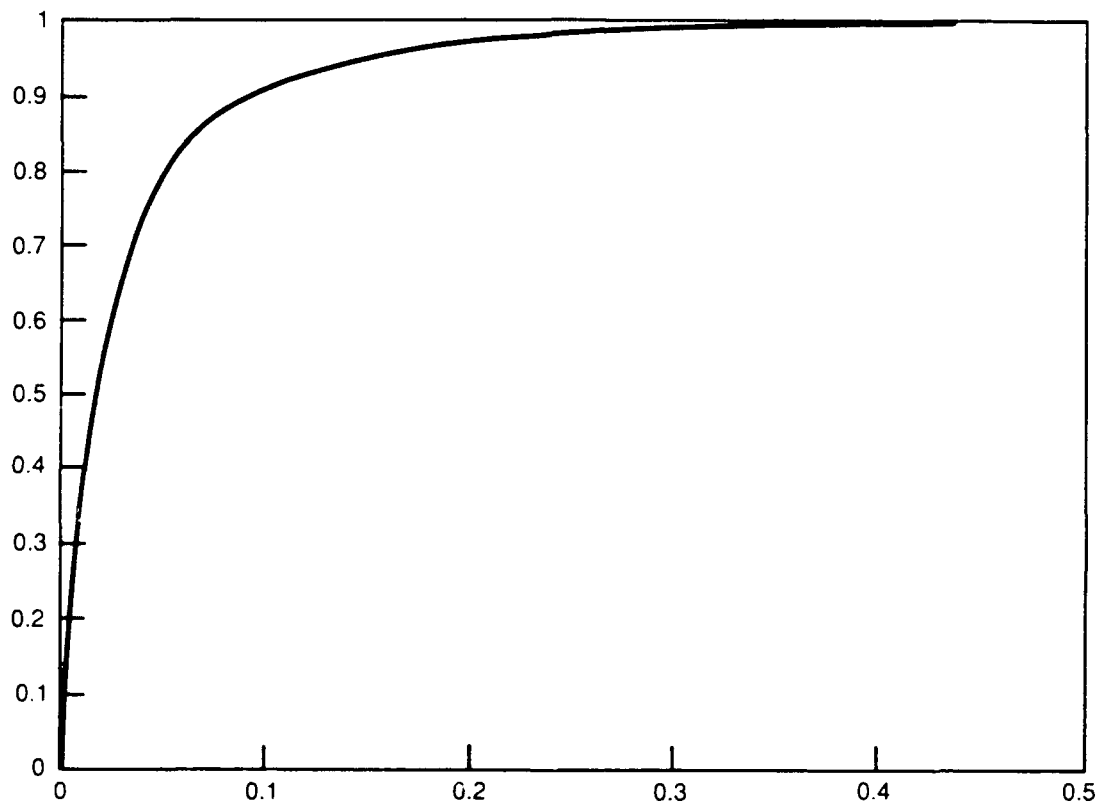
Figure 1. Cumulative spectrum for a 30,000 year run of the Lorenz 27-variable model of climate (see Section 4).

length is small, the statistical reliability of the low frequency estimates is based on few samples and is therefore low.

The low frequency components in a time series can be accentuated by filtering. One means of filtering is to integrate or, for equally spaced data, to form the cumulative sum. If $f(\omega)$ is the estimate of the power spectrum at angular frequency ω of the observed time series $x(t)$, then $f(\omega)/\omega^2$ is the estimate (except at zero frequency) of the spectrum of the cumulative sum $X(n)$

$$X(n) = \sum_{t=1}^{n} x(t).$$

The summing highlights the low frequency components of $x(t)$ without completely eliminating high frequency terms, as would be the case for low pass filtering. In the time domain, long runs of similar values of $x(t)$ will be exhibited as nearly monotonic increases or decreases of $X(t)$, while a series with less correlation will show a much more ragged wave. As will be shown, a measure of the irregularity of the cumulative sum $X(n)$ provides a measure of the persistence within the calculated or observed time series $x(t)$.

In the case where the $x(t)$ are normally distributed <u>and</u> independent, with vanishing mean and unit variance, the running sum $X(n)$ can be described as a classical Brownian process. The increment $X(n + k) - X(n)$ is a normal variable with mean 0 and variance k which is independent of $X(n)$ a<u>nd</u> the values $X(m)$ for $m < n$ since $x(t)$ are by assumption uncorrelated. In a classical Brownian process, there is no memory of the past. The lack of memory is a direct result of the independent nature of $x(t)$. Because $x(t)$ are independent, their index of persistence, denoted by P, and which will be defined shortly, will vanish. For notational purposes we note that the

variance of $X(n)$ scales as n^{2H} where

$$H = 1/2$$

for a Brownian process.

At the opposite extreme to a Brownian process with zero persistence we consider a process in which the increments of $x(t)$ are a constant random number x_1.

$$x(t) = x_1$$

with $E[x_1] = 0$ and $E[x_1^2] = 1$. The cumulative sum for this process is simply

$$X(n) = x_1\,n$$

and the variance of $X(n)$ is given by

$$\text{var}\,[X] = E[x_1^2]\,n^2 = n^2$$

so that the variance scales as n^{2H} where $H = 1$. For this process persistence is perfect, since the run length of identical values equals the length of the record. If, in analogy with the correlation coefficient, we assign $P = 1$ to perfect persistence and $P = 0$ to the absence of persistence, then P is related to the scale factor H by

$$P = 2\,H - 1; \qquad -1 \le P \le 1.$$

Negative values of P correspond to antipersistence.

Hurst (1951), a hydraulic engineer, first introduced the use of cumulative sums to analyze long hydrological records by considering the storage of water in reservoirs. The amount of water in a reservoir at any one time is determined by the cumulative inflow minus the evaporation and draw down. In a managed reservoir, the inflow fluctuates widely but the draw down is

5

more or less constant. Motivated by reservoir considerations, Hurst analyzed a number of records using a rescaled range (R/S) analysis described in detail in the following section. Hurst's results were puzzling. Hydrological records did not scale according to $H = 0.5$ or $P = 0$ but showed values of H that corresponded to positive values of P. For reservoir height, persistence is understandable, since it depends on rainfall in prior seasons as well as rainfall in the current year, and therefore the system has memory. Hurst also found $H > 1/2$ for long temperature records where the origin of "memory" is not as clear.

This paper examines persistence in climate and models of climate. In section 3, we describe the theory of process with stationary increment and introduce the notion of fractional Brownian motion first studied by Kolmogorov (1940) and later by Levy (1953). Such processes have memory and positive persistence. A variety of methods to determine P, including R/S analyses, are employed in Section 4 to determine the persistence of long (26,000 years) records generated by a low order climate model. The results of applying these methods to observational records are discussed in Section 5.

2 RESCALED RANGE (R/S) ANALYSIS OF INDEPENDENT RANDOM VARIABLES

In order to illustrate R/S analysis we first consider a series where $x(t)$ are mutually independent random variables with a common distribution function. For convenience in calculation, the mean of $x(t)$ is set to zero and the values are normalized such that the variance is unity. We form the cumulative sums of $x(t)$ in order to bring out the low frequency behavior

$$X(n) = \sum_{t=1}^{n} x(t).$$

Order statistics are more robust than conventional statistical measures. To apply order statistics we form the range $R(n)$ by setting

$$M(n) = \max [0, X(1), X(2), ..., X(n)]$$
$$m(n) = \min [0, X(1), X(2), ..., X(n)]$$

and

$$R(n) = M(n) - m(n).$$

For ease in interpretation it is useful to modify this definition. Suppose that the total length of the record is N. Instead of considering the sums $X(n)$, we consider instead their deviations from a line drawn joining the origin to the point $(N, X(N))$. Thus we replace the random variables $X(n)$ by

$$X^*(n) = X(n) - nX(N)/N$$

and define the corresponding variables $R^*(n)$, $M^*(n)$ and $m^*(n)$ accordingly. The random variable $R^*(n)$ will be called the adjusted range of the cumulative sum.

In applications, the adjusted sum is to be preferred because it appears to have greater sampling stability (it is more robust) and it eliminates a trend in case that $E[x(t)] \neq 0$. Since the sums $X(N)$ are asymptotically normally distributed, the asymptotic distribution of the range is independent of the distribution of $x(t)$. Summing not only emphasizes the low frequency components of $x(t)$ but also ensures a robust quality to the statistics of $X(n)$, such as the range or the adjusted range.

A straightforward calculation of the range for normally distributed $x(t)$ leads to

$$E[R(n)] = 2(2n/\pi)^{1/2}$$

and

$$\text{var } [R(n)] = 4n(\ln 2 - 2/\pi)$$

(Feller, 1951). The statistics for the adjusted range show the same dependence on the length n over which the sum is taken

$$E[R^*(n)] = (n\pi/2)^{1/2} \sim 1.2533 \, n^{1/2}$$
$$\text{var } [R^*(n)] = (\frac{\pi^2}{6} - \frac{\pi}{2}) \, n \sim 0.07414 \, n.$$

The result that the range varies as the square root of the length of the sum depends critically on the assumption that the variables $x(t)$ are independent. If this assumption does not hold, then $X(n + 1)$ depends on $X(n)$, and since $X(n)$ is determined by the previous n values of $x(t)$, $X(n+1)$ depends on $x(t)$, $t = 1...n$. The cumulative sums provide a measure of past correlation if the dependence of the range, or the adjusted range, does not vary as the square root of the summation interval.

The quantity R/S is formed by dividing the adjusted range of $X(n)$ by $S(n)$, which is the standard deviation of $x(n)$. For stationary Gaussian

8

processes, division by $S(n)$ adds little to the analysis; $x(t)$ can always be normalized to have unit variance. However, for nonstationary processes with long range dependence, division by $S(n)$ leads to a more stable statistic.

Based on the above considerations, we can apply R/S analysis to test for correlation in random number generators commonly employed in computer packages. The random numbers were generated according to the algorithm proposed by Park and Miller (1988), which is under consideration for adoption as a IEEE standard. Figure 2 shows a sequence of 8192 normally distributed variates supposed to be uncorrelated. The observed distribution (see Figure 3) closely approximates the normal distribution. Figure 4 illustrates the cumulative sum of the variables in Figure 2 with the ragged curve typical of a Brownian process. The R/S statistics are obtained by first calculating the adjusted range R^* for the total record of length $N = 8192$ and normalizing by the standard deviation of the record of length N. The process is repeated for chunks of record of length $N/2, N/4, ...N/1024$, and the resulting value of the scaled adjusted range, R/S, is obtained by averaging over the $2, 4, ...1024$ non-overlapping intervals. The smallest interval contains eight points. The concept of range begins to break down at small interval and chunks smaller than 8 are not used. A plot of the log (R/S) against log (n), where n is the length of the interval over which R/S is estimated, determines the scaling exponent H in the relation

$$R/S \sim n^H.$$

For the series shown in Figure 2 the scaling slope is 0.51 (see Figure 5). For $n = 8192$, the difference in adjusted range for slopes 0.5 and 0.51 is 11.8, while the standard deviation of the estimate of adjusted range is 24.09, so that the value 0.51 is not statistically distinct from 0.5. For comparison, a slope of 0.6 would produce an adjusted range more than eight standard

9
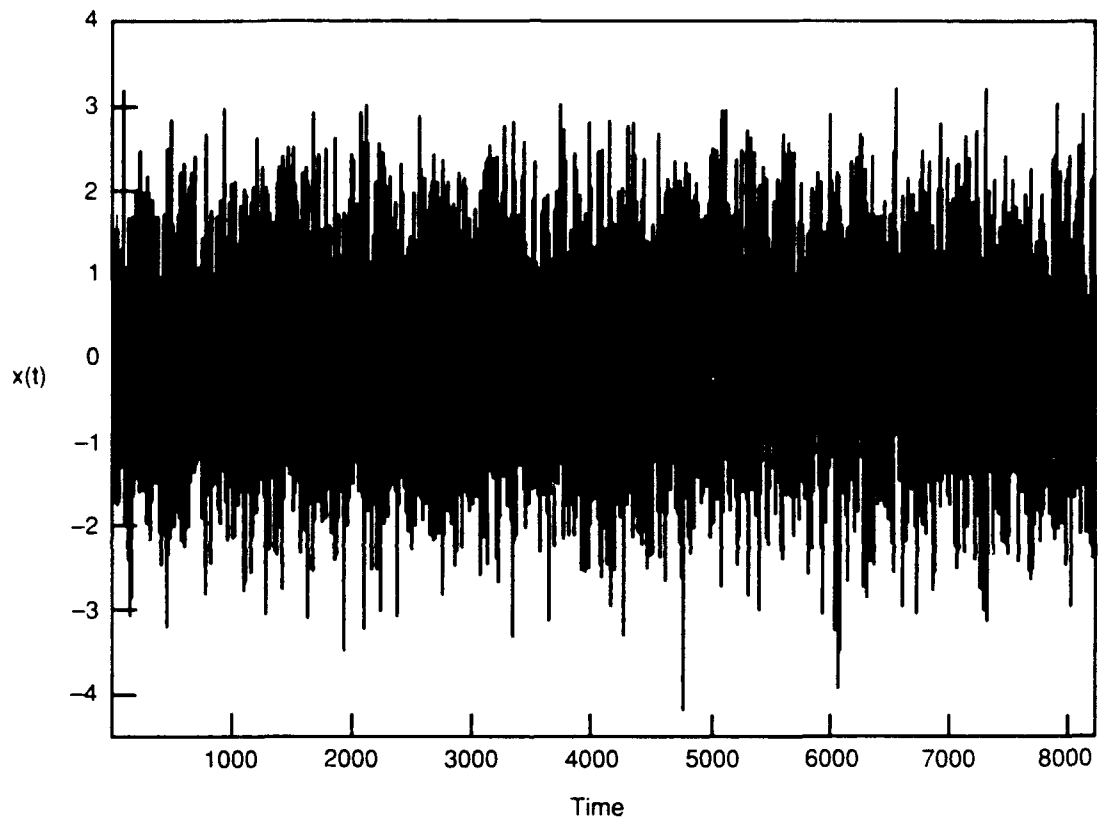
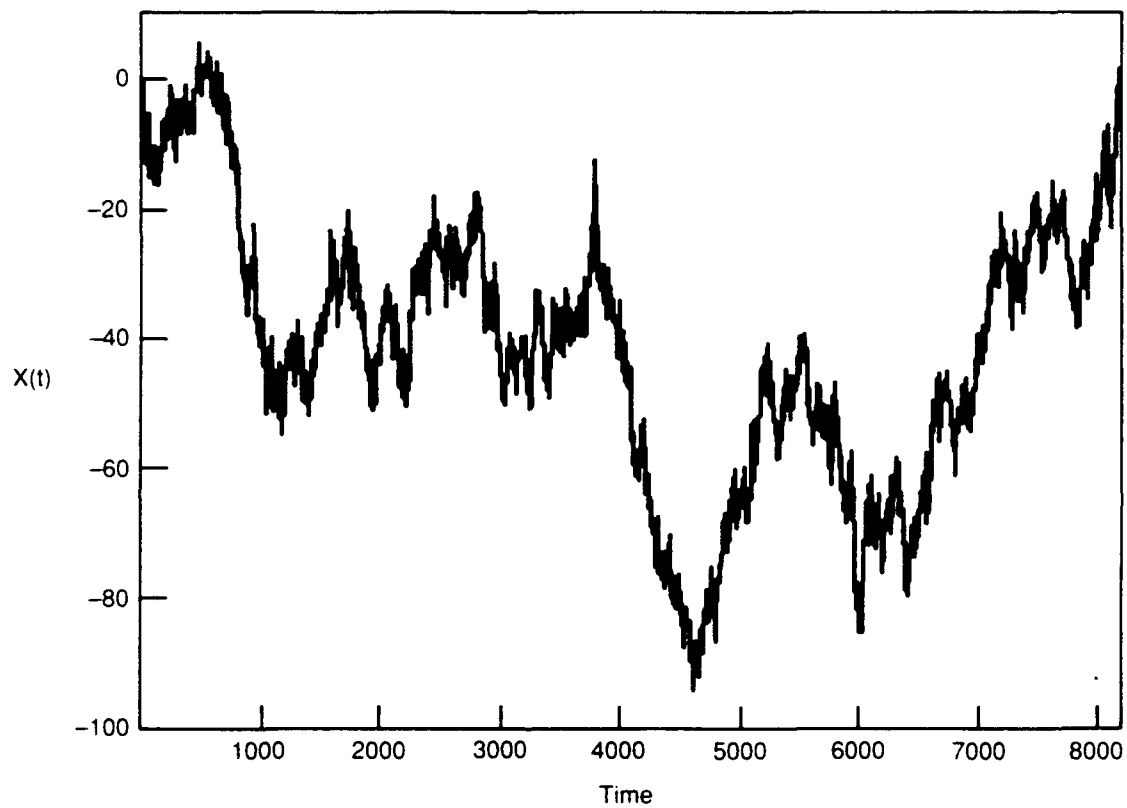**Figure 2.** Time sequence of 8192 normally distributed random variables with zero mean and unit variance.
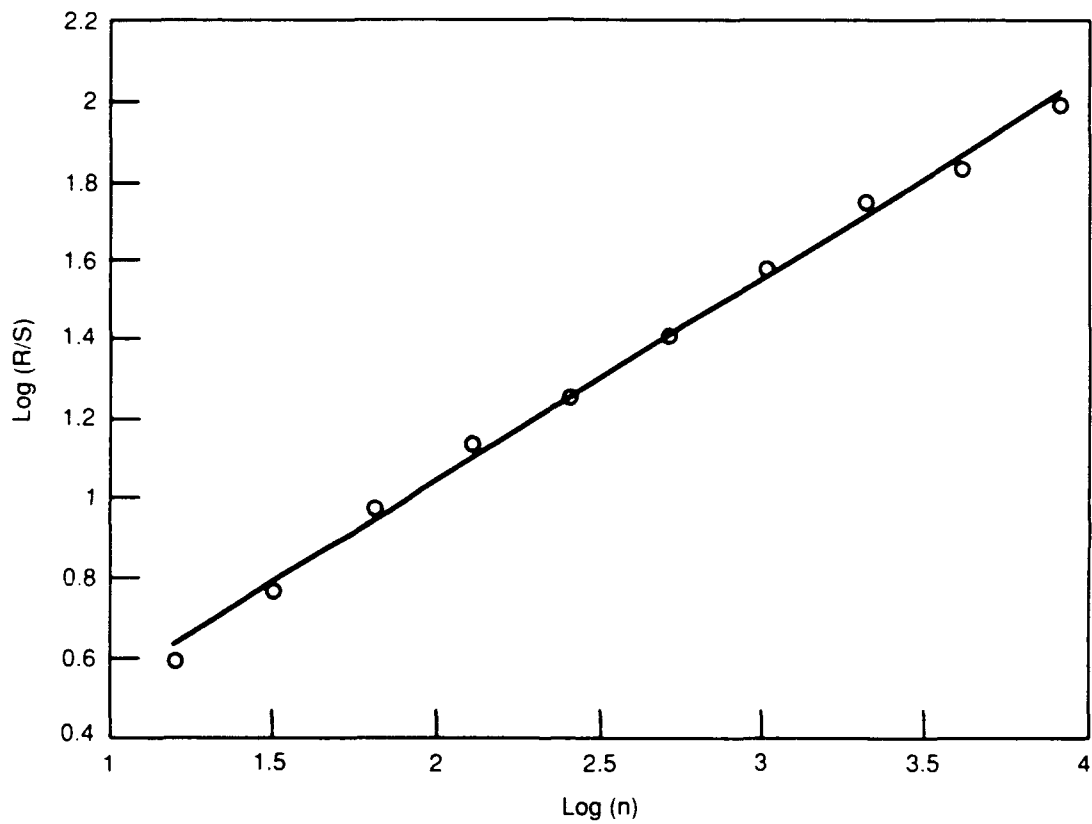
10
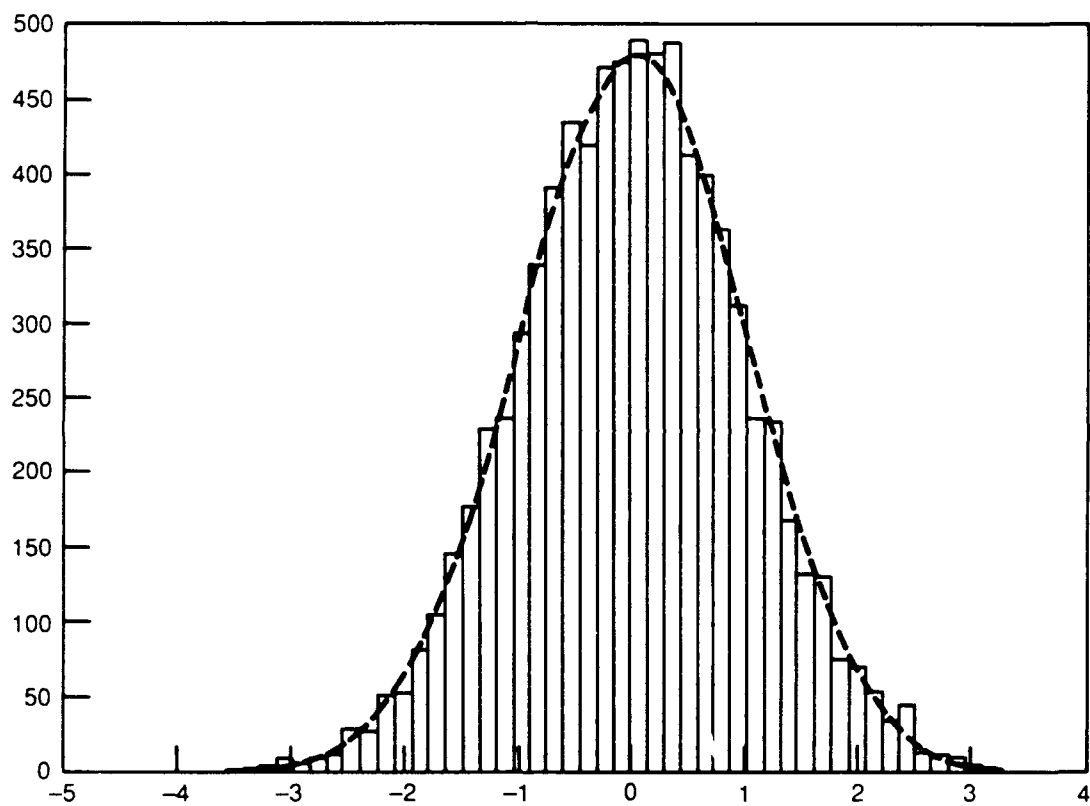
Figure 3. Histogram of the x values shown in Figure 2.

11

Figure 4. Cumulative sums X(t) of x(t) shown in Figure 2.

12

Figure 5. Dependence of the log of the R/S statistic on the log of the interval over which R/S is estimated for the variable shown in Figure 2. Slope of curve is 0.5136.

deviations from the expected value for a random independent variable.

We note that while the range for the cumulative sum of the normally distributed independent variables scales with length of observation as $n^{1/2}$, the range of the elements in the sum $x(t)$ increases far more slowly, essentially as $(\log n)^{1/2}$ (Leadbetter et al., 1980; Abarbanel et al., 1991).

A trend introduces a long term persistence into a record, as would a pure harmonic, $A \sin(2\pi f + \phi)$. We take the frequency f and amplitude A as fixed but the phase ϕ is taken as random having a uniform distribution. For $x(t)$ equal to a sine wave, the range of the cumulative sum $X(t)$ in the limit of large n is $A/\pi f$ while the limit of the standard deviation is $A/2$, so that the limit of R/S is $2/\pi f$. The statistic R/S for a pure sine wave does not vary according to $n^{1/2}$, so it shows persistence (actually perfect antipersistence).

The situation differs in the limit of large n when white Gaussian noise of zero mean and unit variance is added to the process

$$x(t) = A\sin(2\pi ft + \phi) + g(t).$$

The cumulative sum $X(n)$ then satisfies the inequality

$$G(n) - A/2\pi f < X(n) < G(n) + A/2\pi f$$

where $G(n)$ is the cumulative sum of the noise $g(t)$.

In the limit of large n, the ranges of $X(n)$ and $G(n)$ are identical since the range of $X(n)$ varies as $n^{1/2}$. The variance of the sine wave plus noise in the limit of large n is

$$E[S(n)] = 1 + A/2$$

and therefore the statistic R/S has an expected value for large n of

$$E[R/S] = 2\left(\frac{2n}{\pi f(1 + A/2)}\right)^{1/2}.$$

The value of R/S, as well as the time required to reach the limit, is determined by $A/2\pi f$. Very low frequency sine waves will lead to persistence in the presence of uncorrelated noise. Lines in the spectrum at low frequencies will affect the estimate of R/S, as will a linear trend. High frequencies will not show up in the R/S statistic.

The existence of long term correlations can have important consequences. For example, the physical model for Brownian motion is one of a small particle being bombarded by molecules in thermal equilibrium. The $n^{1/2}$ dependence flows from the assumption that the molecules in thermal equilibrium have uncorrelated motions. With vanishing correlation the ordinary diffusion equation applies. Statistical independence at large time and/or space scales is an essential ingredient of the concept of thermal equilibrium; the existence of longer term correlations would invalidate the assumption of thermal equilibrium.

3 FRACTIONAL BROWNIAN MOTION AND PERSISTENCE

For the moment we return to the cumulative sum of independent, stationary increments, and form the structure function for $X(k)$, $X(k) = \sum_{t=1}^{n} x(t)$ defined by

$$D(k) = E[(X(k) - X(0))^2]$$

where by convention we can set $X(0) = 0$. The structure function for $X(n + m)$ is

$$
\begin{aligned}
D(n + m) &= E[(X(n + m) - X(0))^2] = E[(X(n + m) - X(m))^2] \\
&+ E[(X(m) - X(0))^2].
\end{aligned}
$$

But we also have

$$E[(X(n + m) - X(m))^2] = D(n)$$

<u>if</u> the increments are stationary <u>and</u> independent and

$$D(n + m) = D(n) + D(m) \text{ for } n > 0, \ m > 0 \ .$$

The only possible functional form for the structure function is

$$D(n) = Cn,$$

identical to the functional form for the variance of the range described in the previous section.

The spectral density $F(\omega)$ associated with $X(t)$ is related to the continuous structure function $D(\tau)$ by

$$D(\tau) = 4 \int_0^\infty (1 - \cos \omega \tau) F(\omega) \, d\omega \ .$$

The spectrum of the independent, random increments is a constant

$$f(\omega) = \frac{C}{2\pi},$$

therefore the spectrum of the cumulative sum of independent, random increments is

$$F(\omega) = \frac{C}{2\pi\omega^2}.$$

Kolmogorov (1940) generalized the concept of Brownian processes to consider structure functions of the form

$$D(\tau) = C\tau^{2H}$$

with $0 < H < 1$. The limits on H come from examining the spectral representation of $D(\tau)$. Since from dimensional considerations

$$F(\omega) \sim 1/\omega^{2H+1}$$

and since $1 - \cos \omega\tau$ varies as ω^2 near zero frequency, the spectral representation for $D(\tau)$ exists if $2H + 1 < 3$, which fixes the upper limit for H. At high frequencies, an ultraviolet catastrophe can only be avoided if $H > 0$. These limits on H allow the persistence index $P = 2H - 1$ to run from -1 to 1. Since the spectrum for X can be written as

$$F(\omega) = \frac{C}{\omega^{2H+1}},$$

the proportionality constant C_1 is related to C in the relation for the structure function by

$$C_1 = \frac{C}{4 \int_o^1 (1 - \cos x)\, x^{-2H-1} dx}$$

$$= \frac{\Gamma(2H + 1)\sin(2H\pi/2)C}{2\pi}.$$

18

The power law dependence of the structure function implies that D and X are self similar. Thus a change in scale in the time variable

$$n = hn$$

leads to a scale change in X

$$X(n) = a(h)X(hn) .$$

For classical Brownian motion we have

$$a = h^{-1/2}$$

or more generally

$$a = h^{-H}$$

and correspondingly the variance scales as

$$\text{var}[X(n)] = h^{2H} \text{ var } [X(hn)] .$$

The self similar character of the cumulative sums implies that the "box dimension" of the curve is not unity. The box dimension is determined by the relation between the number of boxes required to cover the curve and the dimension of the individual boxes. If the length of the record is N, then N/hn segments of length hn are needed to cover the time axis. In each segment the range of the record is $R(hn) = h^H R(n)$ and we need a stack of $h^H R(n)/ha$ boxes of height ha to cover the range. The total number of boxes is then

$$B(h; a, n) = \frac{h^H R(n)}{ha} \frac{T}{hn} \sim h^{H-2} \sim h^{-D_B}$$

where D_B is defined as the box dimension. In terms of the parameter H the box dimension is

$$D_B = 2 - H.$$

The determination of the box dimension for the sequence of random variable shown in Figure 2 is illustrated in Figure 6. Here the box dimension $D_B = 1.48$, which yields an H value of 0.52; this can be compared to the value of 0.51 obtained by an R/S analysis.

Series with values of the parameter $H > 1/2$, or $P > 0$, show persistence; that is, the process generating the time series has memory. Levy (1953) demonstrated this by showing that for $X(t)$ generated as a Brownian-like process, $X(t)$ can be represented as

$$X(t) = \frac{1}{\Gamma(\frac{2H+1}{2})} \int_0^t (t-s)^{\frac{2H-1}{2}} \, dW(s)$$

where $dW(s)$ is the increment of an ordinary Brownian process ($H = 1/2$). This representation clearly demonstrates the long term memory associated with a fractional Brownian process.

Persistence of a fractional Brownian process can also be shown by examining the correlation properties. In particular, the past increments are correlated with future increments. Given the increment $X(0) - X(-n)$ from time n to time 0, the probability of having an increment $X(n) - X(0)$ averaged over the distribution of past increments is proportional to
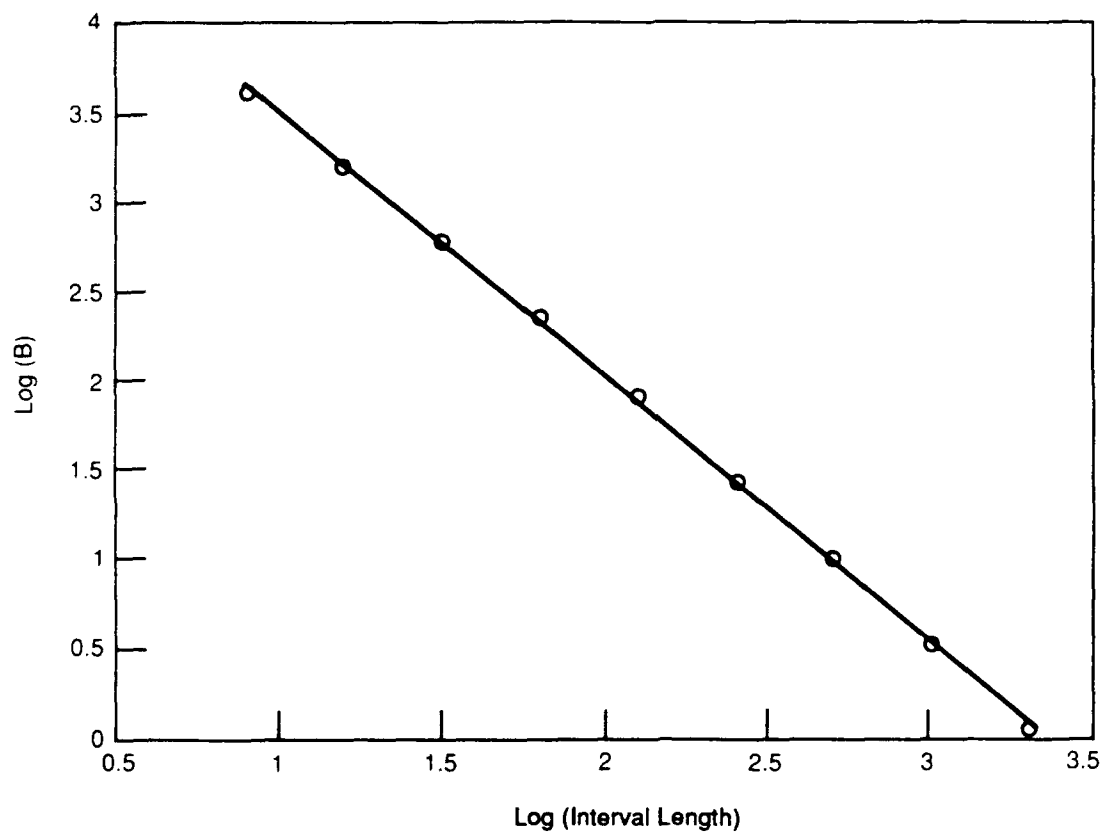
$$E[(X(0) - X(-n))(X(n) - X(0))].$$

For convenience we choose the origin so that $X(0) = 0$. The correlation function of future increments with past increments is then

$$C(n) = \frac{E[-X(-n)X(n)]}{E[X(n)^2]}$$

where we have normalized with the variance. The above scaling relations $E[(X(n) - X(k)^2] \sim |n - k|^{2H}$ lead to

$$C(n) = 2^{2H-1} - 1 = 2^P - 1.$$

20

Figure 6. Determination of box dimension of the cumulative sum curve shown in Figure 4 for a zero mean unit variance random variable having a normal distribution. Slope = −1.4789, which corresponds to H = 0.5211.

The correlation of future increments with past increments vanishes for strict Brownian motion. However, for positive persistence, the correlation is positive and independent of the time. This implies an increasing trend in the past will produce an increasing trend in the future. In fact, in time limited systems, fractional Brownian motion is an approximate model over some range of time scales but not necessarily for the entire record.

Fractional Brownian motion is illustrated in Figures 7 and 8. In Figure 7, the curve corresponds to a persistence of 0.8. The long runs of $x(t)$ with the same sign show up clearly as nearly straight climbs or descents in the cumulative sum $X(n)$. Antipersistence is shown in Figure 8, giving rise to a highly ragged curve.
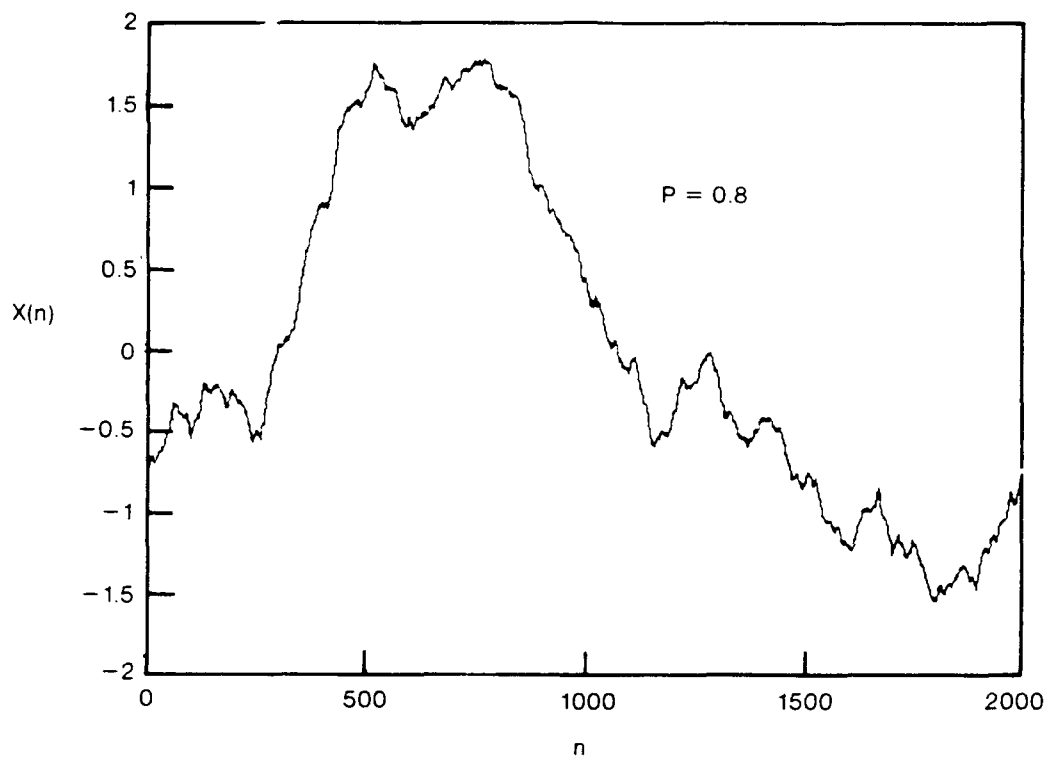
Figure 7. Fractional Brownian motion, P = 0.8

23
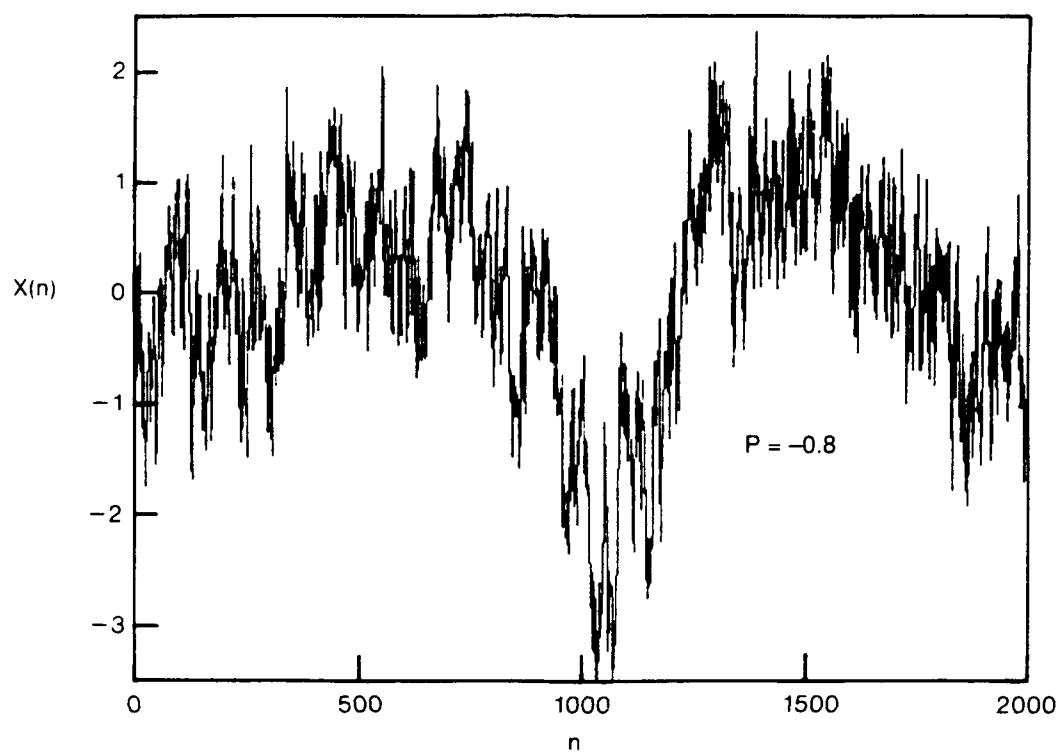
Figure 8. Fractional Brownian motion, P = -0.8

4 ANALYSIS OF LONG-TIME INTEGRA-TION OF LORENZ 27-VARIABLE MODEL

4.1 Introduction

We use values generated by a Lorenz 27-variable model of the atmosphere to illustrate the analysis described in Sections 2 and 3. Lorenz (1984) formulated a "low order" model of atmospheric change with a goal of exploring a "moist" atmosphere. A moist atmosphere contains bulk liquid water and water vapor that can form clouds. A shallow ocean provides water for evaporation as well as a heat sink or source. Clouds form or dissipate in response to changes in relative humidity with consequent changes in albedo and the radiative balance.

In the model, the albedo is proportional to the cloud cover; since freezing is excluded from the model there are no changes in albedo due to varying snow and ice cover. The cloud cover is parameterized in terms of the relative humidity. Locally, the relative humidity drops with higher temperature and the albedo decreases, leading to more solar radiation to heat the earth. The cloud-albedo feedback allows for significant swings in local and global average temperature.

The radiative and thermodynamic processes that characterize a moist atmosphere introduce complicated nonlinear terms into the governing equations. In order to limit the computational complexity, Lorenz (1984) introduced a number of simplifying approximations. Lorenz found that for short

term integrations the model produced results that qualitatively were in reasonable agreement with observed atmospheric behavior and concluded that the model was suitable for production runs. Abarbanel et al. (1990) and Abarbanel et al. (1991b) investigated the climate of the Lorenz model by integrations over several thousand years. We have continued these studies using a 25,521 year integration with a 1.25 hour time slip and averaging over the year and over the surface of the model to obtain "global" annual average temperature. The model does not contain either a seasonal or diurnal cycle.

An important point to note is that the model contains only one long time constant: the thermal time constant for warming the shallow ocean, about ten years. The ocean is taken to have ten times the heat capacity of the atmosphere. Horizontal heat flow in the ocean is not allowed. Thus there is a permanent equator to pole temperature gradient that undergoes small fluctuations in response to the poleward transfer of heat by the atmosphere. The pole to equator thermal gradient forms a large scale coherent structure with a time scale equal to the length of the record.

4.2   Statistics of the Lorenz-27 Model

The results of a 25,521 year integration of Lorenz-27 for global annual average temperature are displayed in Figure 9. The temperatures have been scaled to unit variance with a standard deviation of 0.364°K (see Table 1). The time history of the temperature resembles a white noise process and any long term trends or persistence are masked by the irregular short term behavior. The histogram for the temperature values is shown in Figure 10. The temperature values closely approximate a normal distribution. The skew-
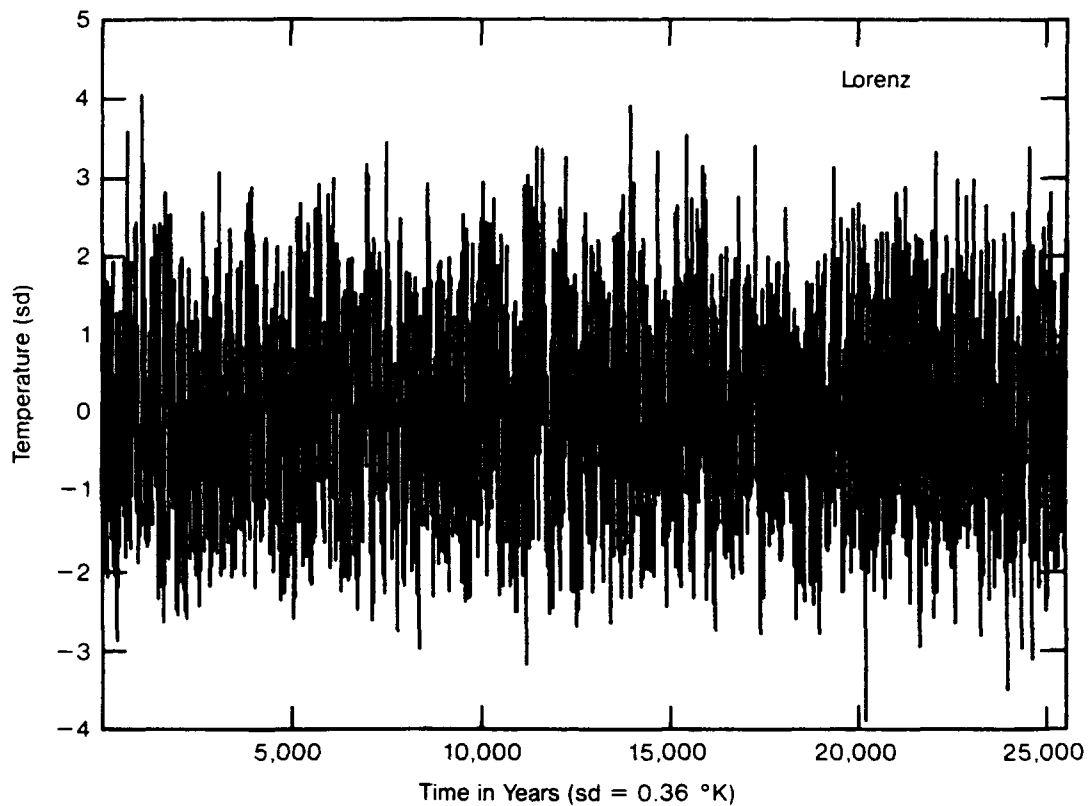
Figure 9. Variation in global annual average temperature for a 25,521 year integration of the Lorenz 27-variable model. The values are reduced to zero mean and unit variance. The standard deviation of the raw temperature record is 0.364°K.
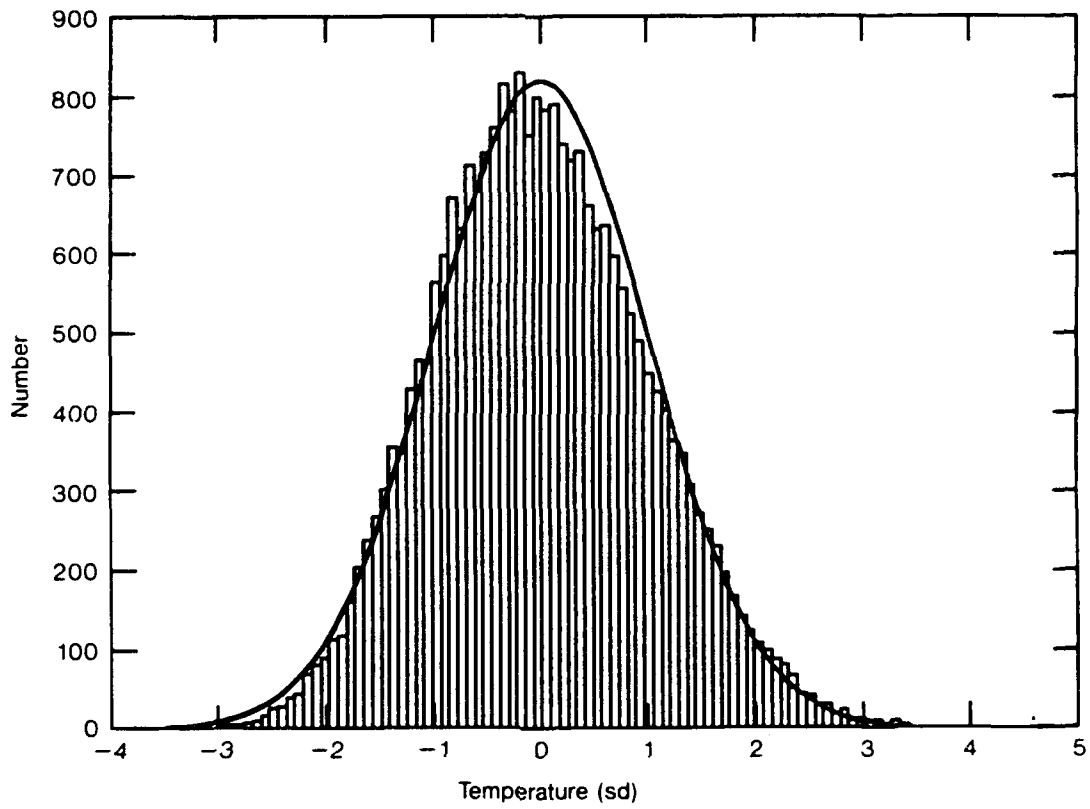
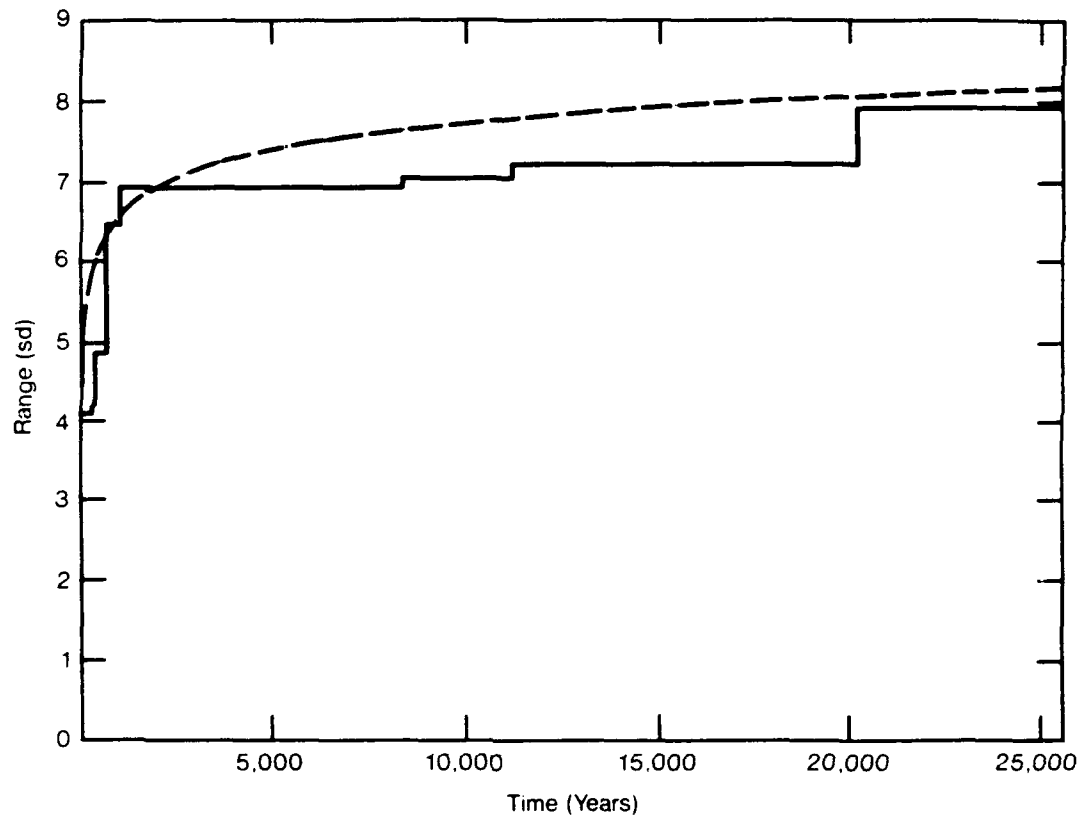**Figure 10.** Histogram of the temperature values shown in Figure 9.

Table 1

Standard Deviation of Global Annual Average Temperature of Long (25,521 years) Integration of Lorenz 27-Variable Model

Interval (Years)	Standard Deviation (Degrees K)
1 - 25,521	0.3.64
50001 - 15 000	0.374
15001 - 25000	0.359
40003 - 4302	0.270
23401 - 23700	0.321
12374 - 12673	0.342
14001 - 14140	0.341

ness coefficient is 0.18, which is not significant. The distribution also closely approximates a normal distribution in the tails. The range of the temperature in units of standard deviation is shown in Figure 11 together with the expected value for a variable drawn from a normal population. The auto-correlation function for the temperature record is shown in Figures 12 and 13. The correlation shows an exponential decay with a first zero at about 65 years. At longer lags the correlation function shows structure, though the paucity of samples for long lags raises question about statistical significance.

The power spectrum of Lorenz-27 shows abundant energy at low frequencies (see Figures 14 and 15). The spectrum shows a number of peaks at periods greater than 100 years. The cumulative spectrum (Figure 1) establishes that 90 percent of the variance of the record displayed in Figure 9 is contained at frequencies less than 0.1 cycles per year (cpy). For periods less than 16 years, the spectrum drops off as $f^{-2.2}$ (see Figure 16).

Figure 11. Variation in range of the temperature values shown in Figure 9. The dashed curve represents the expected value of the range for a normally distributed random variable. (Abarbanel et al. 1991a)
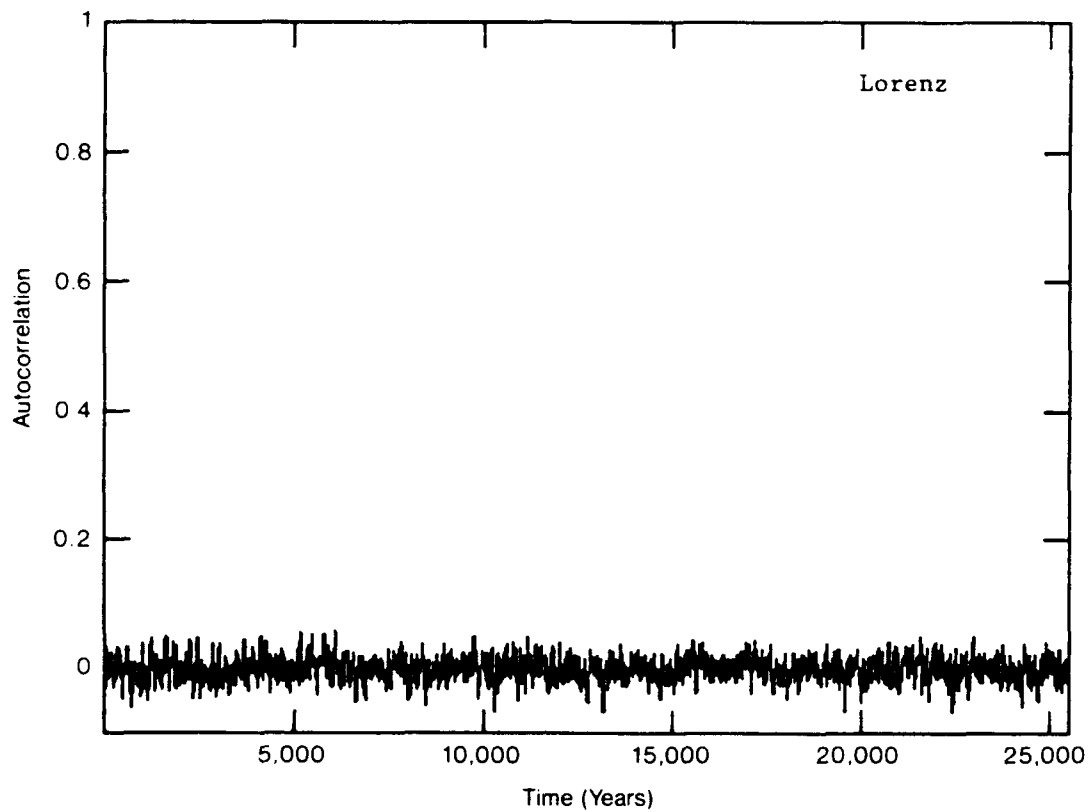
Figure 12. The autocorrelation function for the time series shown in Figure 9. The correlation function obtained by taking the Fourier transform of the power spectrum is shown in Figure 14.
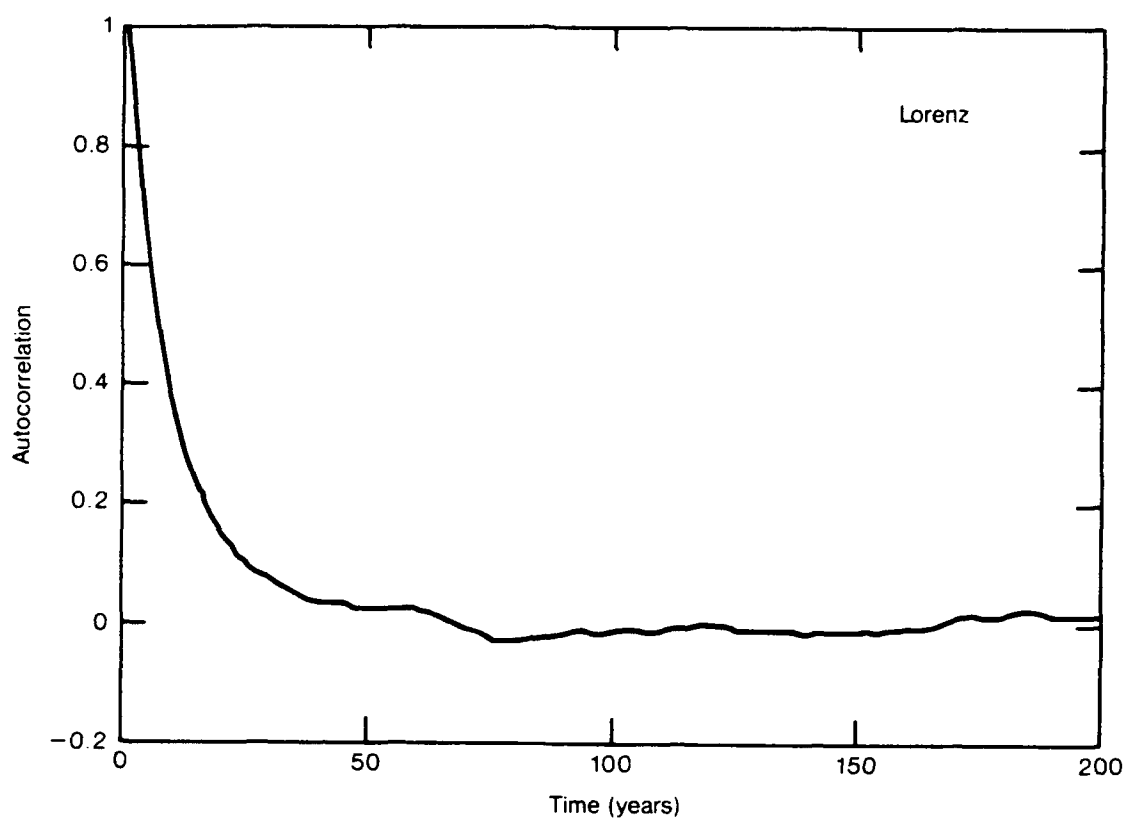
Figure 13. Autocorrelation at small lags for the record shown in Figure 9.
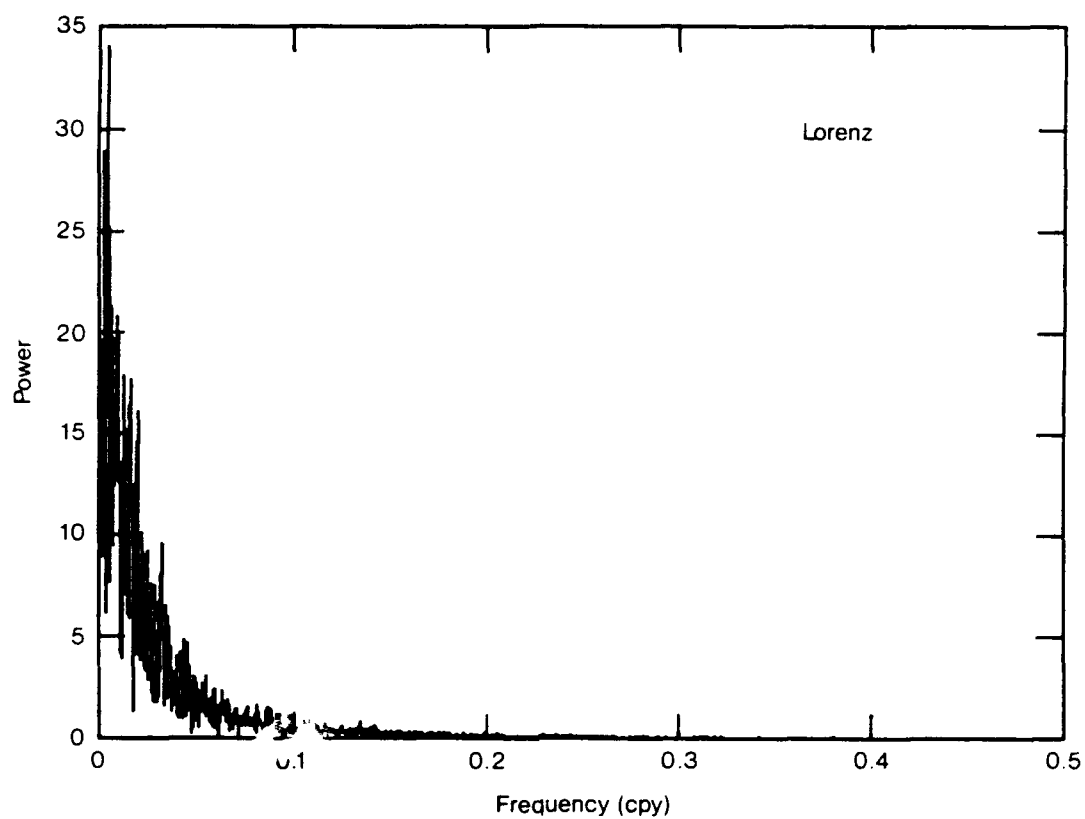
Figure 14. Power spectrum of the record displayed in Figure 9. The frequency is measured in cycles per year (cpy). The units for the power are $(^\circ K)^2/cpy$.

33
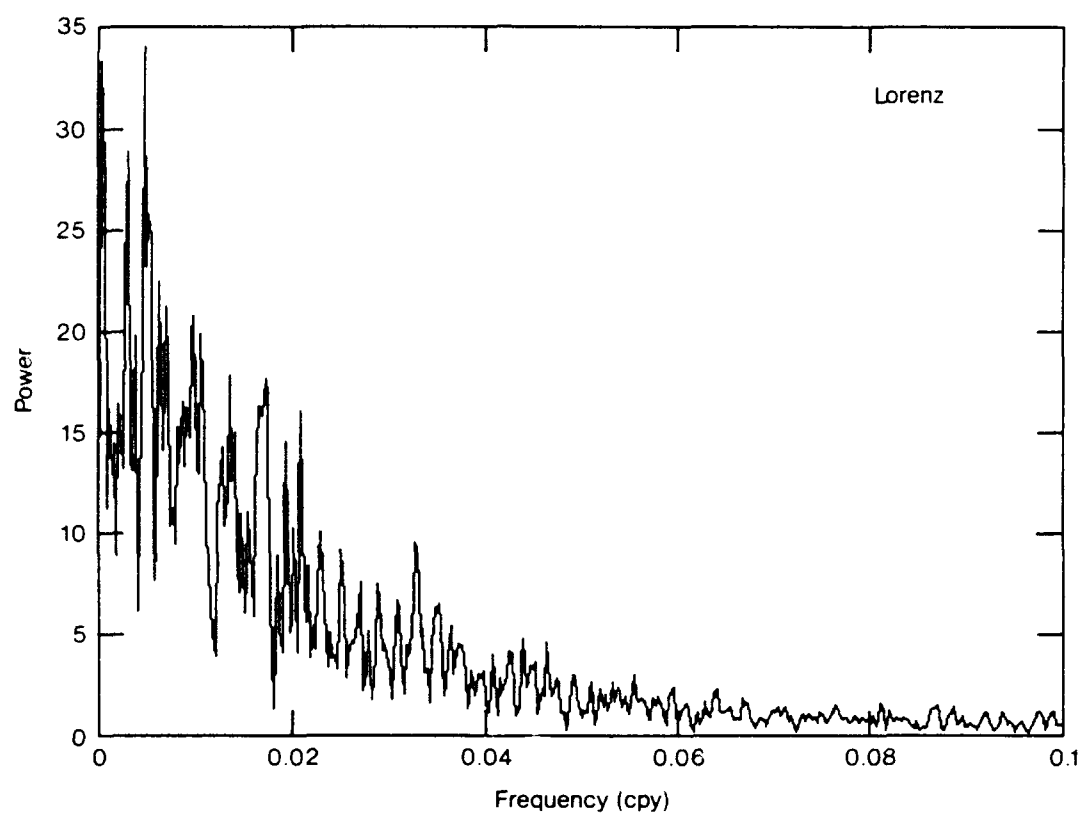
Figure 15. The low frequency portion of the power spectrum shown in Figure 14.
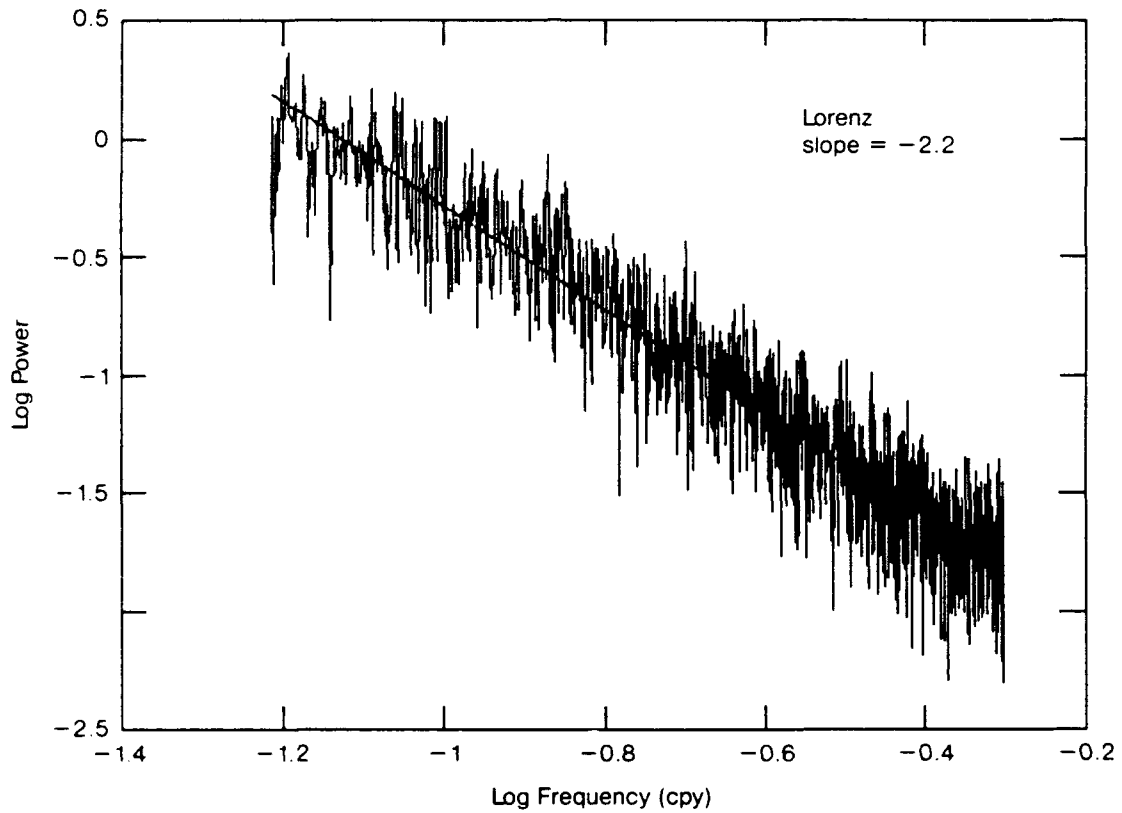
Figure 16. The high frequency portion of the spectrum given in Figure 14 in a log-log representation. The best least squares line fit to the spectrum is $f^{2.2}$ suggesting that for time scales of 16 years or less, the record given in Figure 9 can be modeled as a Brownian motion with independent increments.

35

Thus at high frequencies the record resembles a classical Brownian walk derived from white noise. The behavior deviates greatly from this model at low frequencies, which contain the bulk of the variance.

4.3 R/S Analyses of Lorenz-27 Model

An R/S analysis of the record shown in **Figure 9** yields a persistence index $P = 0.34$ for intervals greater than about 20 years (see Figure 17). An alternative interpretation of the analysis is shown in Figure 18. In this interpretation there are three scales for the variation in temperature. At high frequencies, periods smaller than 20 years, the variations can be modeled as classical Brownian motion with independent increments. For periods between 20 years and 400 years, the record has high persistence, $P = 0.64$. At still longer intervals (400 years to 25,000 years), the record shows weak persistence, $P = 0.07$. Whatever correlation exists at intermediate time scales is lost at long (> 400 years) time scales.

In terms of the spectrum of the record, in the frequency interval corresponding to time intervals between 400 and 20 years, the power spectrum should drop off as f^{-P}. The observed drop off yields an estimate for P of 0.71 (see Figure 19) as opposed to the estimate of 0.64 derived from the R/S analyses. The spectral estimate of P depends sensitively on the highest frequency used in obtaining the slope, since at high frequencies the slope of the spectrum is -2.2 (see Figure 16). Persistence is closely tied to the character of the spectrum at low frequencies. The higher the fraction of the variance that is carried by the low frequencies, the more likely that a record will show persistence. In order to distinguish persistence due to low frequency lines in

Figure 17. Results of an R/S analysis of the record given in Figure 9. The logarithm of the scaled range is plotted against the logarithm of the interval over which the rescaled range was estimated.
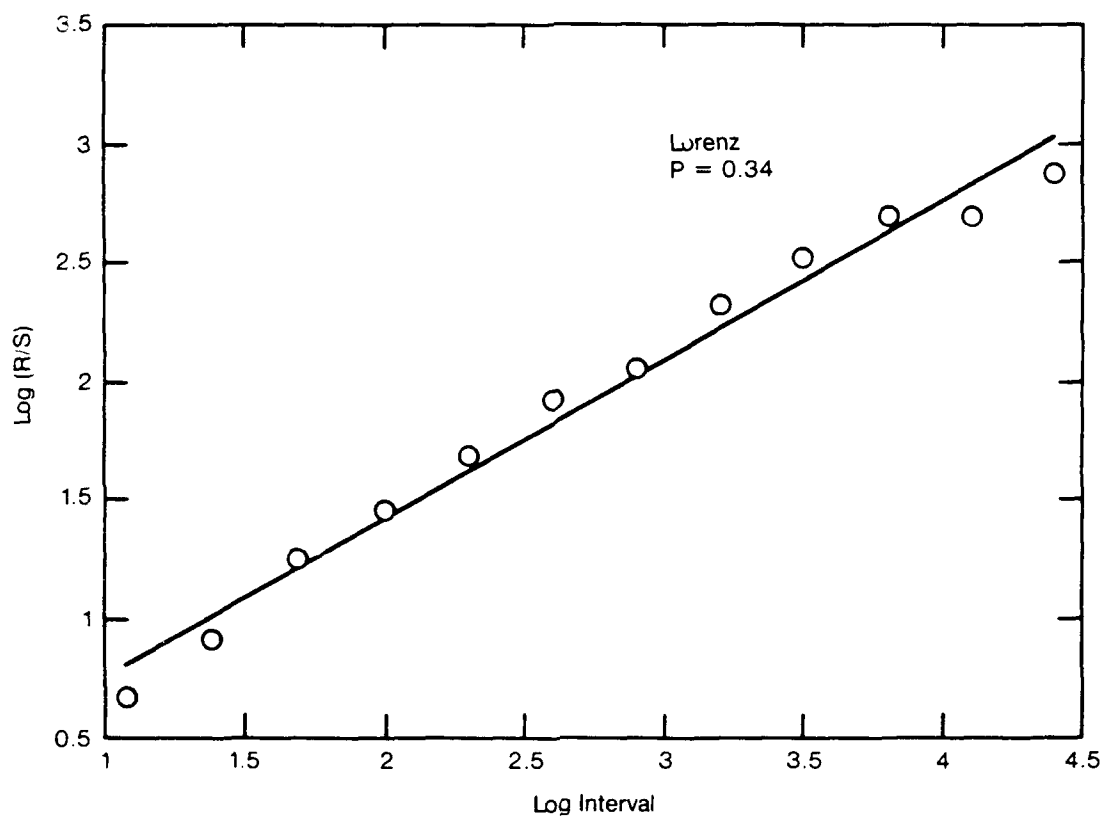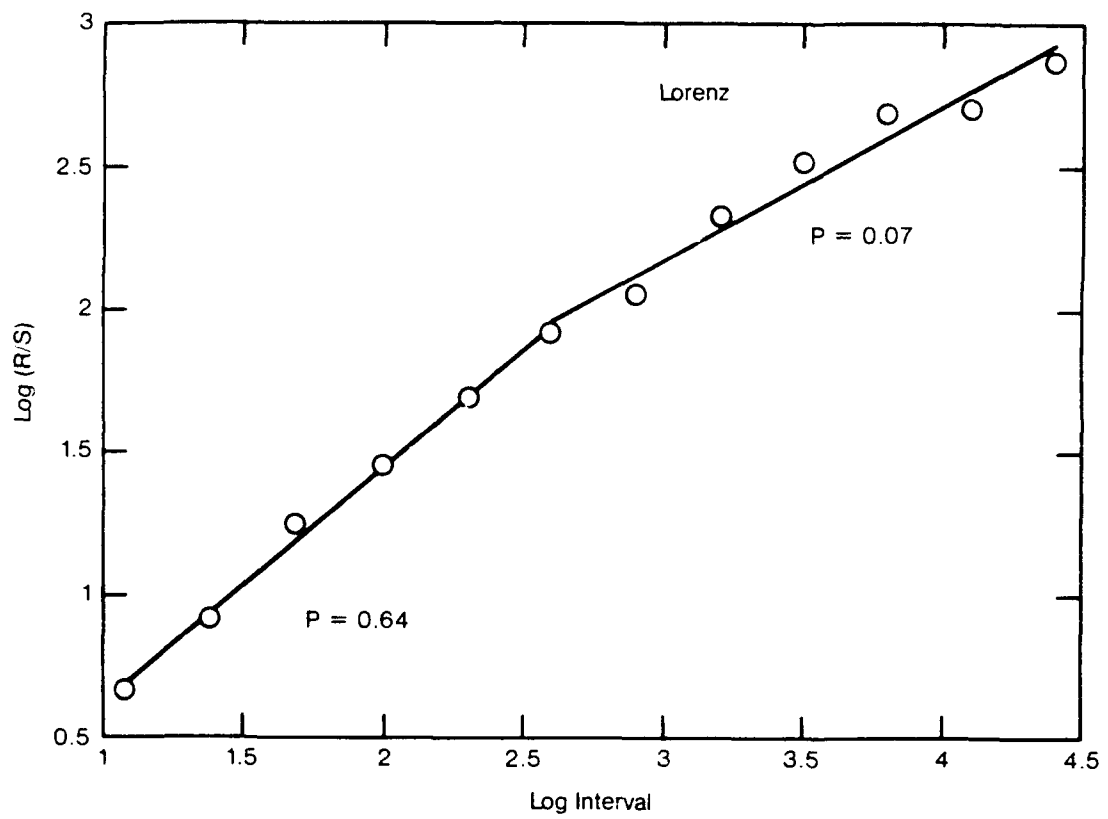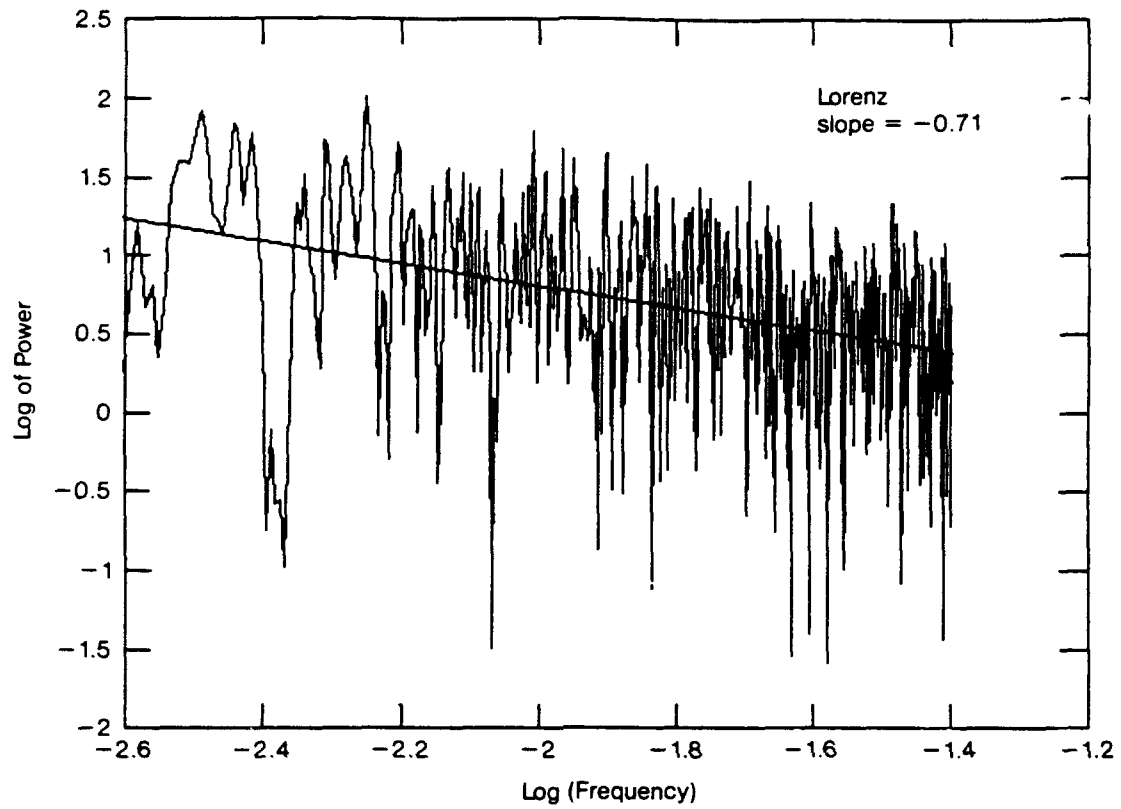
37

Figure 18. Alternative intepretation ot the R/S analyses of the record shown in Figure 9.

Figure 19. Log-log plot of the spectrum of the record given in Figure 9. The derived persistence is 0.71.

the spectrum from persistence due to a continuous low frequency spectrum a much longer record would be required (see Section 2).

The Lorenz-27 climate model shows three time intervals with differing behavior. At high frequencies, the model behaves as an oscillator excited by white noise. At an intermittent range of frequencies, the variations in global annual average temperature show partial coherence as measured by the persistence index. Over these time scales one may be able to make predictions about average behavior provided the averages are longer than 20 years. At still longer periods (greater than 400 years) partial coherence is largely lost and attempts at prediction might be expected to fail.

4.4 Analysis of Persistence in Short Records

The estimates of persistence discussed in Section 4.3 were based on the analysis of the entire record. In climate studies good global annual average temperature are very much limited in length, to about 110 to 140 years, though there is a simple station record covering some 300 years (Abarbanel et al. 1991b). Table 2 lists R/S estimates of persistence for shorter intervals of the record. The higher values of persistence shown for randomly selected records of length 140 and 300 years are due to persistence for time intervals of 20 to 400 years. The shorter records do not sample intervals greater than 400 years and thus do not pick up the lack of persistence for intervals longer than 400 years.

Table 2

**Persistence in Lorenz-27 Model as Estimated
from Intervals of Various Lengths**

Interval	Length (Years)	Persistence All Intervals	20-400 Years
1 - 25521	25521	0.34	0.64
5001 - 15000	10000	0.42	0.74
12374 - 12673	300	0.56	-
4003 - 4302	300	0.61	-
23401 - 23700	300	0.50	-
14001 - 14140	140	0.48	-

4.5 Alternative Methods of Estimating Persistence

The R/S method of estimating persistence appears robust since it depends on use of order statistics (David, 1981; Huber, 1972). Two alternative methods to R/S statistics are the analysis of the variation of the variance with length of the interval and the estimation of the box dimension (see Section 3). In the variance method, the variance of the cumulative sum is calculated for the whole record and then for successive smaller non overlapping intervals obtained by repeated halving of the length of the intervals. The dependence of the variance on the length of the interval then determines $P + 1$, where P is the persistence index.

Table 3 provides comparison of various estimate of the index of persistence. The R/S estimate and the estimate of the box dimension of the cumulative sum of the record are close. The estimate obtained by examining the dependence of the variance on interval length consistently gives a lower

Table 3

**Comparison of Various Methods
of Estimating Persistence**

Record	R/S	Variance	Box Dimension
Lorenz 1 - 25521	0.34	0.22	0.31
Lorenz 5001-15000	0.42	0.30	0.44
Random 1-8192	0.02	-0.02	0.01

value of P than that provided by R/S analysis or the box dimensions.

4.6 Comments on Estimating Persistence

The principal requirement for estimating persistence is a long record. As discussed above, the value of P depends on the length of the record, particularly if persistence becomes small for intervals longer than some interval. The value of P obtained for short records should be viewed with caution. The R/S estimate exhibits greater statistical stability than does the variance estimate.

42

5 PERSISTENCE IN GLOBAL ANNUAL AVERAGE TEMPERATURE

5.1 Observed Temperature Records

The record of global annual average temperature prepared by Jones et al. (1986 a,b) and brought up to date by Jones (1988) and Jones and Wigley (1990) is shown in Figure 20. The record combines observations taken both at land stations and ships at sea. The records are fraught with problems related to the homogeneity of land and marine data, but represent the best information available. The values earlier than 1890 are undoubtedly less reliable than those in subsequent years.
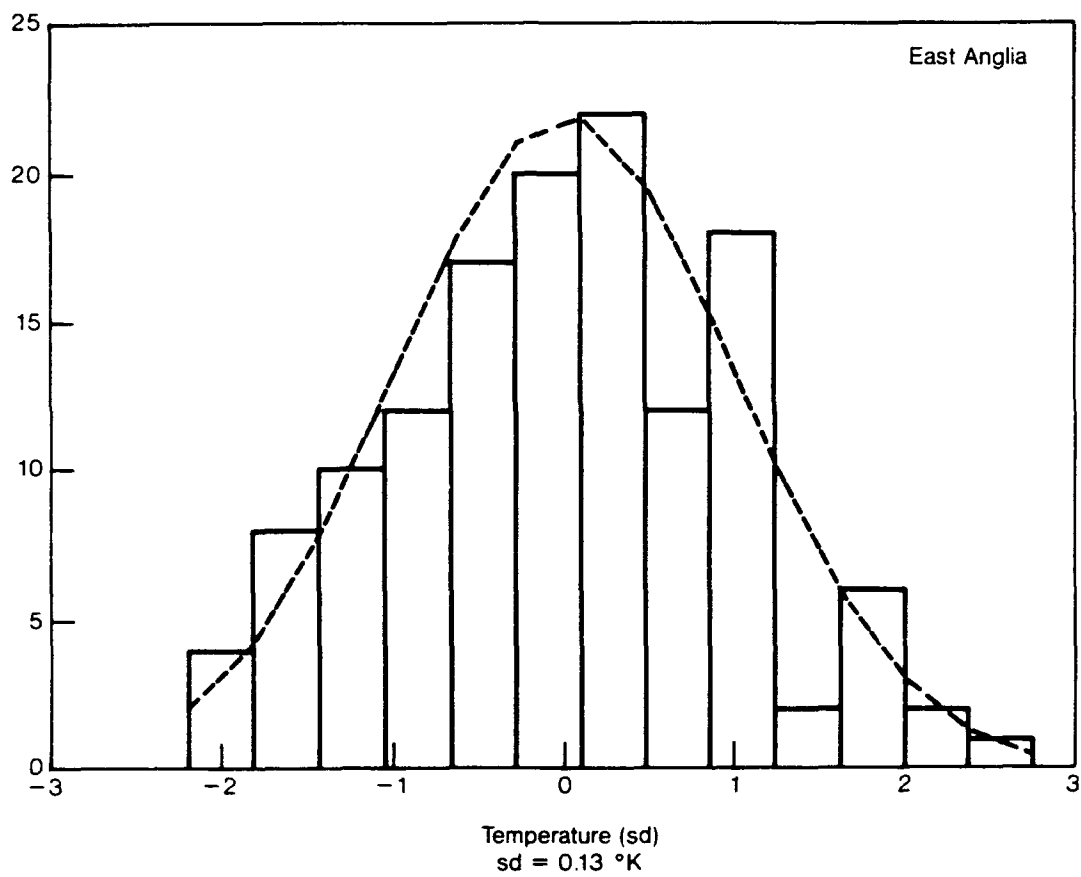
The temperature time series in Figure 20 shows a trend with warming apparent, particularly in the decade of the 1980s. A least squares fit of a linear trend provides a slope of $0.028°K/year$. If the linear trend is removed the remaining time series has a histogram that approximates a normal distribution (see Figure 21). The standard deviation for the series including the trend is $0.17°K$; with the trend removed, the standard deviation is $0.13°K$. The variations of the range of temperature with time provide a measure of the nature of the tails of the distribution. As illustrated in Figure 22, the observed range approximates the expected value of the range of a normally distributed random variable.

Figure 23 shows the power spectrum of the global annual average temperature. The power is concentrated at the lower frequencies with 60 percent
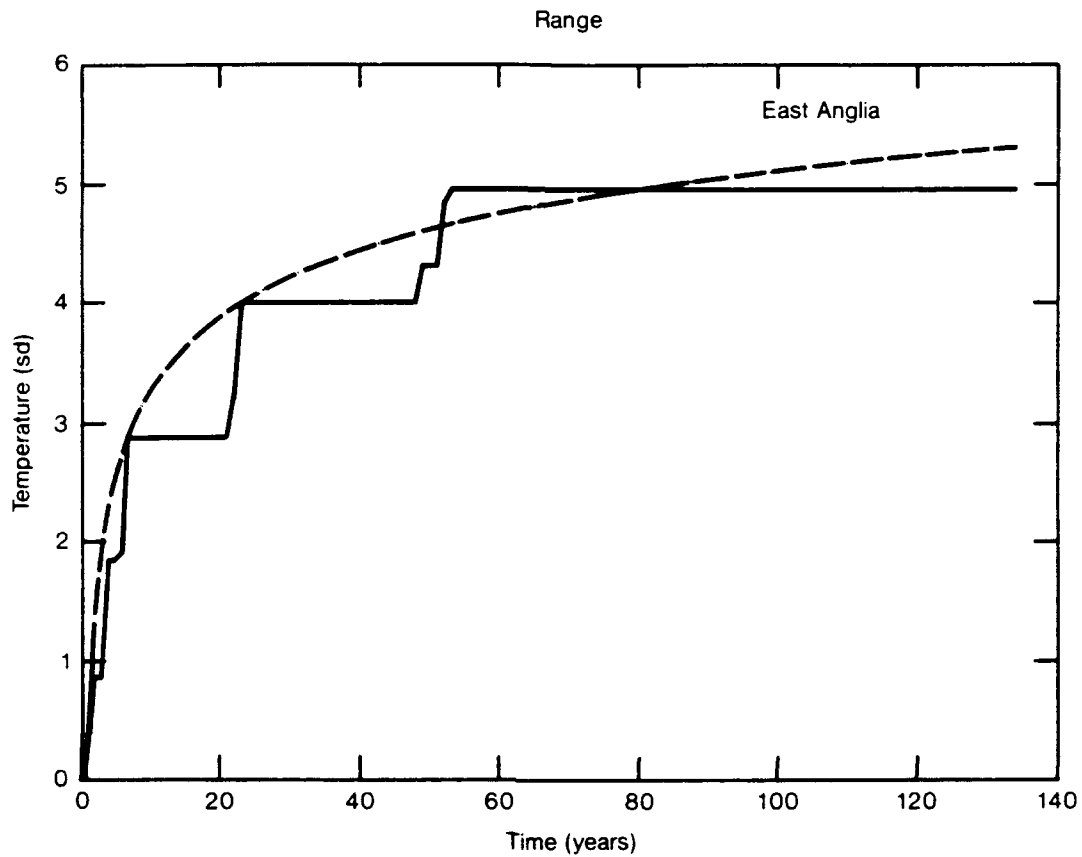
43
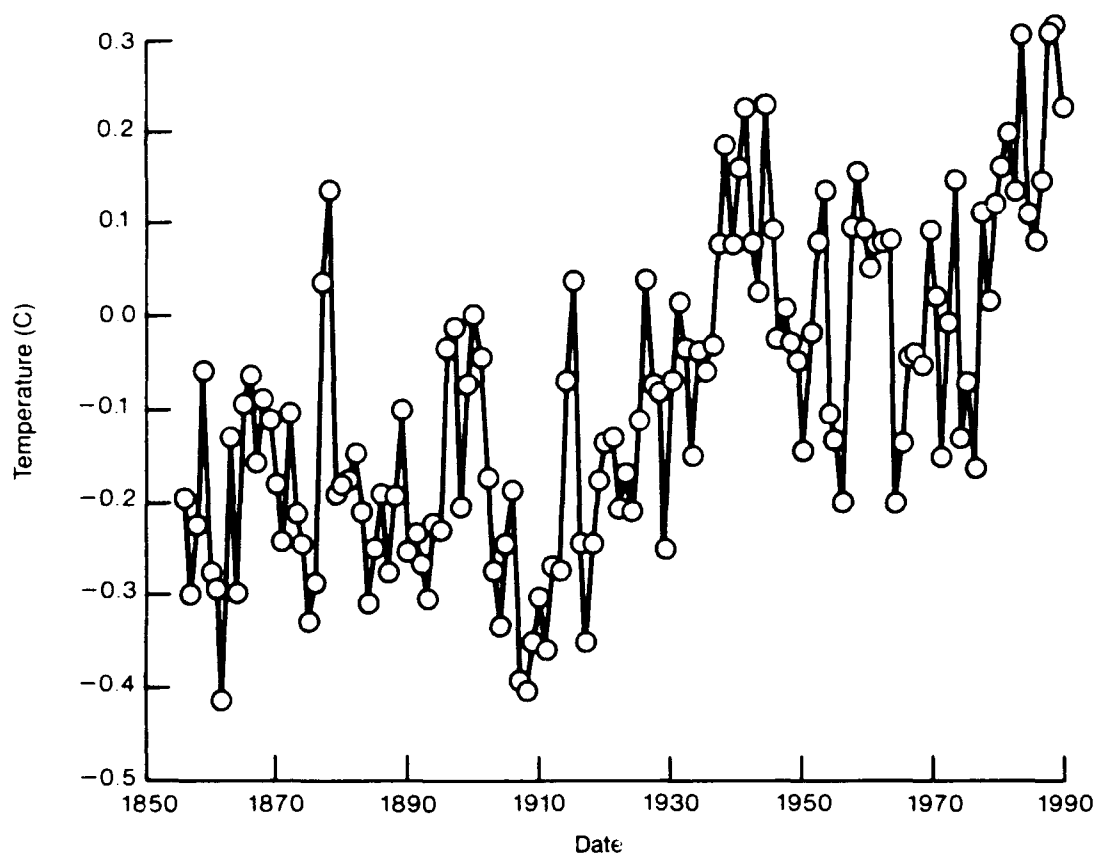
Figure 20. Global annual average surface temperature after Jones et al. (1986 a,b).

44

Figure 21. Histogram for the values of temperature shown in Figure 20 with a linear trend removed. The standard deviation for the record is 0.13°K.

45

Figure 22. Variation of the range of temperature time series of Figure 20 from which a linear trend has been removed. The dashed curve represents the expected value of the range for a normally distributed random variate with zero mean and unit variance.
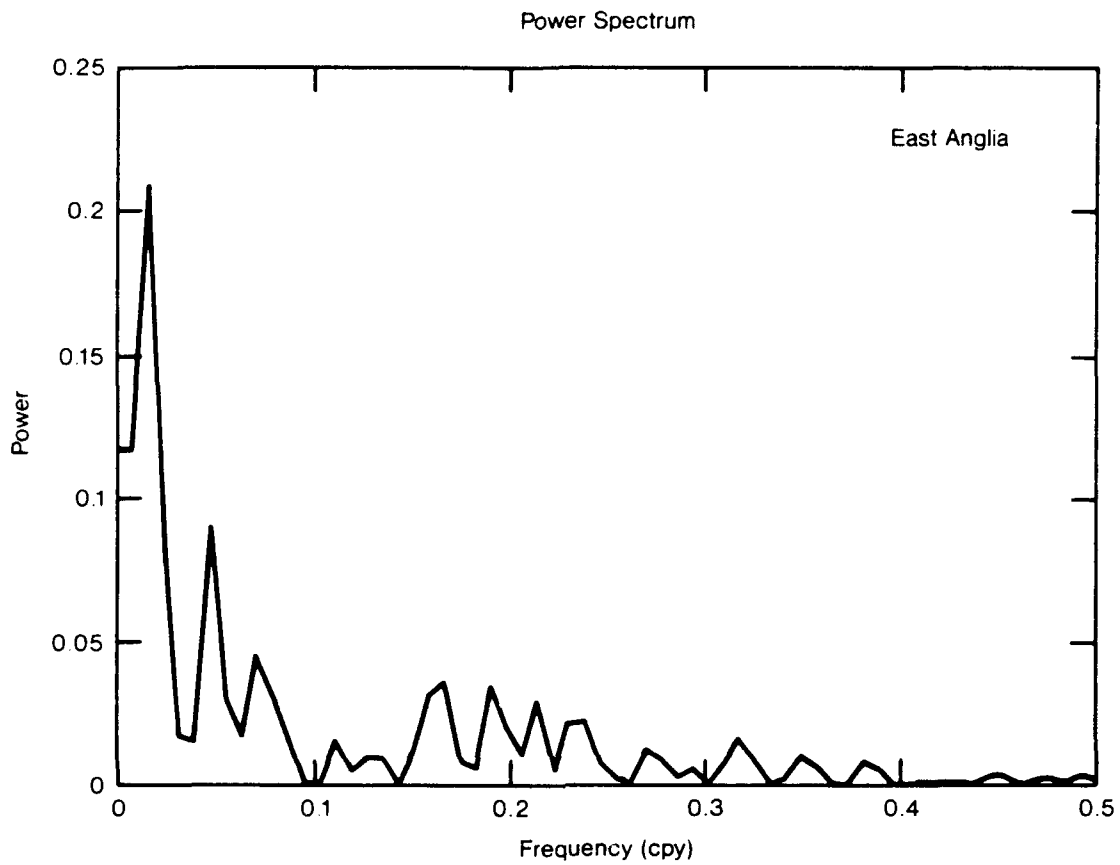
Figure 23. Power spectrum of record shown in Figure 20 after removal of a linear trend.

Table 4

Estimates of Persistence Obtained for
Global Annual Average Temperature Records

| Record	Length (Years)	Standard Deviation (°K)	Index of Persistence
Hansen and Lebedeff (1988)	108	0.22	0.88
same with linear trend removed	108	0.14	0.71
Jones et al. (1986b)	134	0.17	0.78
same with linear trend removed	134	0.13	0.56
Manley (1974)	318	0.68	0.46
same with linear trend removed	318	0.61	0.38

of the variance frequencies lower than 0.1 cycles per year (see Figure 24). The fraction of low frequency variance is less than that displayed in the Lorenz model (90 percent), but the observed record certainly contains uncorrelated measurement error that may well reduce the fraction of energy due to low frequency natural variability.

The persistence index obtained from an R/S analysis of the record shown in Figure 20 is 0.78. If the trend is removed the persistence drops to 0.56 (see Figure 25). The persistence is comparable to that shown in the Lorenz model for short intervals (see Table 2). Table 4 gives the value of the persistence as given by R/S analysis of two other long term temperature records. Hansen and Lebedeff (1988) reduced temperature observations for

48
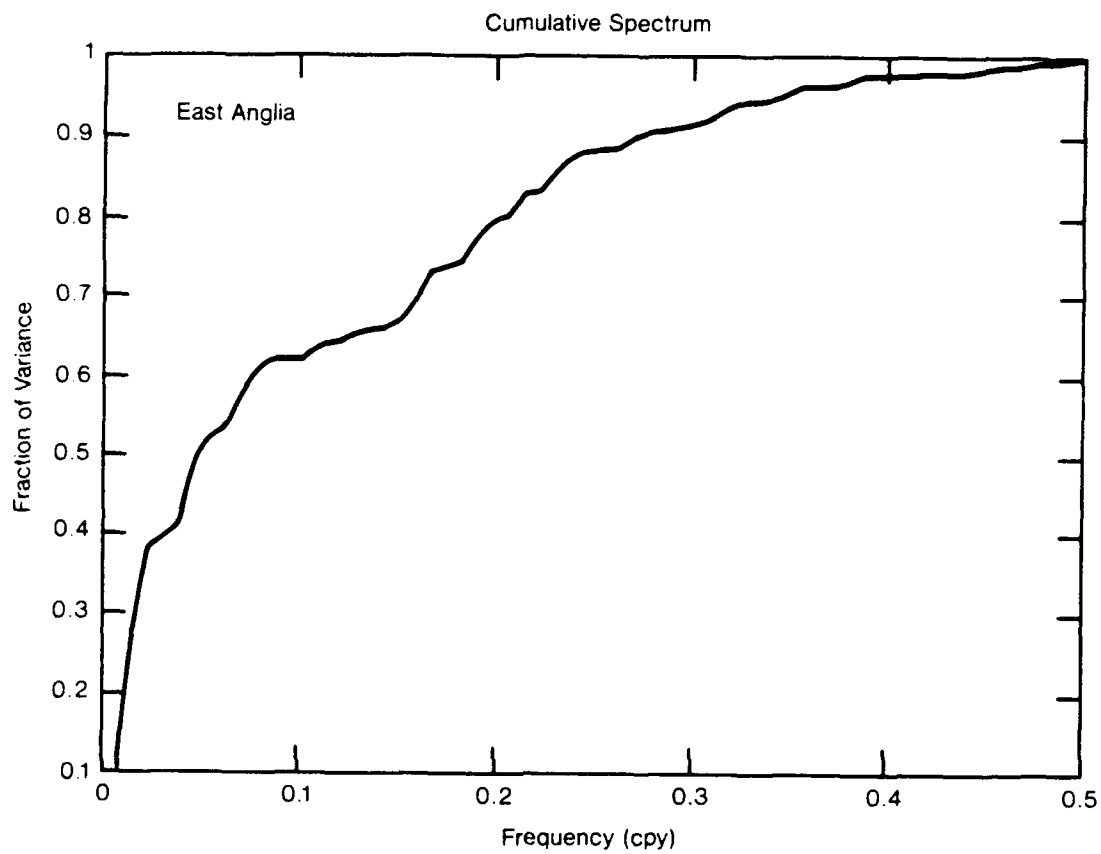
Figure 24. Cumulative spectrum of the time series given in Figure 20.
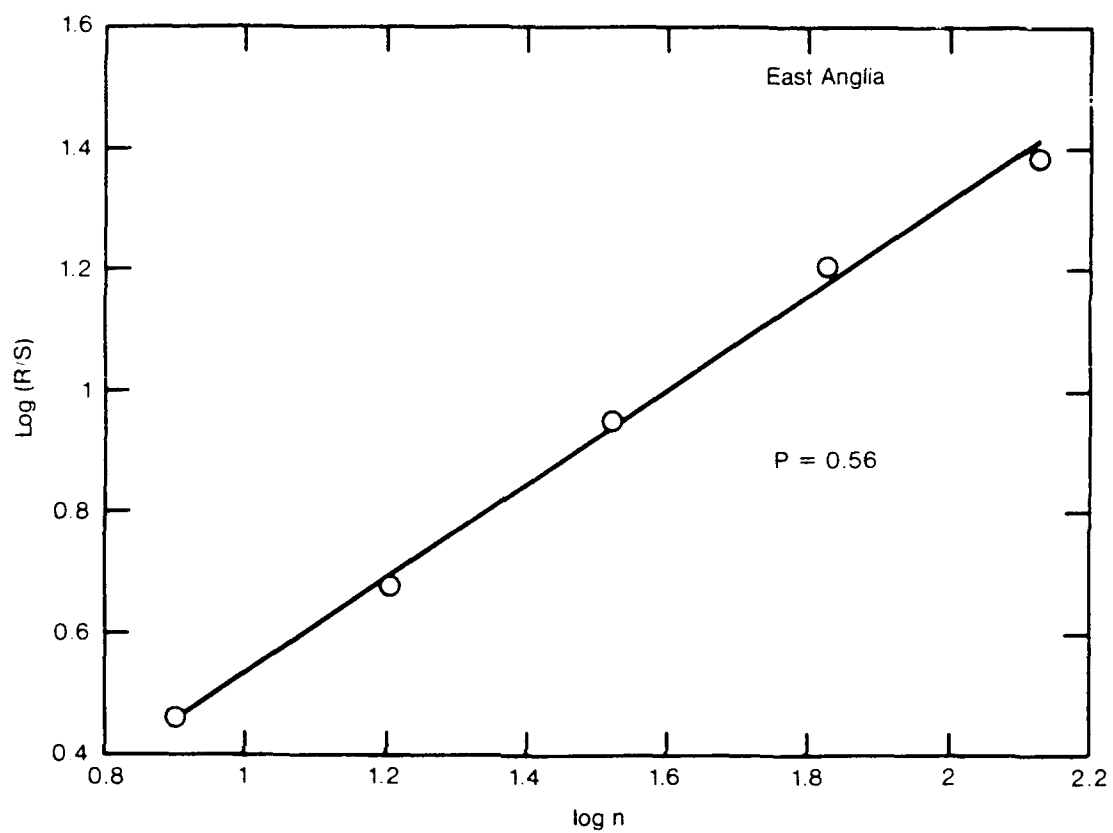
Figure 25. R/S analysis of temperature record given in Figure 20 with a linear trend removed.

land stations only to obtain a global annual mean surface temperature. The Manley (1974) record is pieced together from observations taken at various stations in Central England and is the longest available instrumented temperature record. All the records show evidence for significant persistence. The lower persistence fc the Manley record is probably due to the much higher noise level associated with a single station and with primitive observations during the early part of the record.

## 5.2 Persistence in a Global Circulation Model (GCM)

Through the efforts of Robert Chervin we had available a 101 year run of the Community Climate Model-1 (CCM-1), a widely used general circulation model (GCM). This model contains far greater complexity than does the Lorenz 27-variable model. Clouds, topography and snow and ice feedback are all included. The oceans exchange heat and moisture with the atmosphere but the temperature of the ocean varies according to the seasonal climatological means. There are no ocean currents and, like the Lorenz model, there is no horizontal heat transport within the ocean. The resulting global annual mean for the model is shown in Figure 26. The standard deviation of $0.043°K$ is small compared with the observed records and with the Lorenz model (see Table 4).

The gross statistics of CCM-1 are indicated in Figures 27 and 28. About the mean, CCM-1 approximates a normally distributed variate but the observed range is significantly greater than that expected for a normally distributed variate. The power spectrum for the model results differs dramatically from the observed records and the spectra of the Lorenz model

51
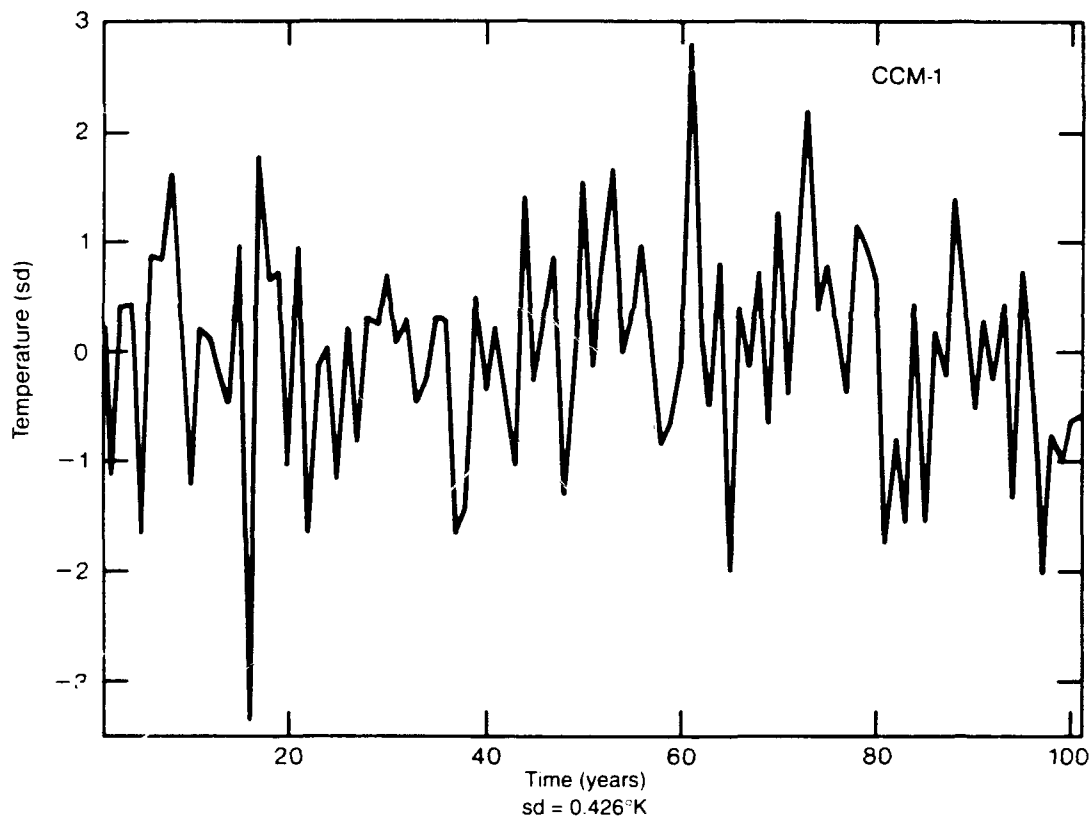
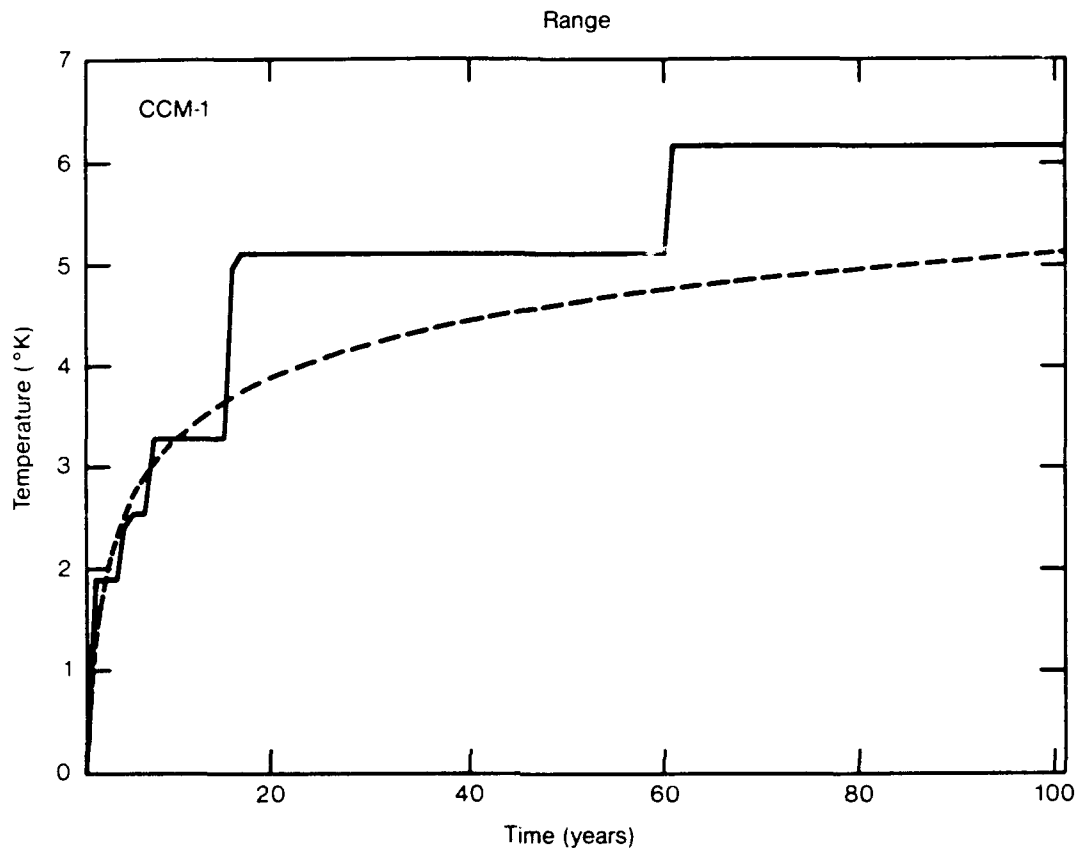**Figure 26.** Global annual average temperature for a 101 year record produced by Community Climate Model-1. (Chervin, 1991 personal communication).

52
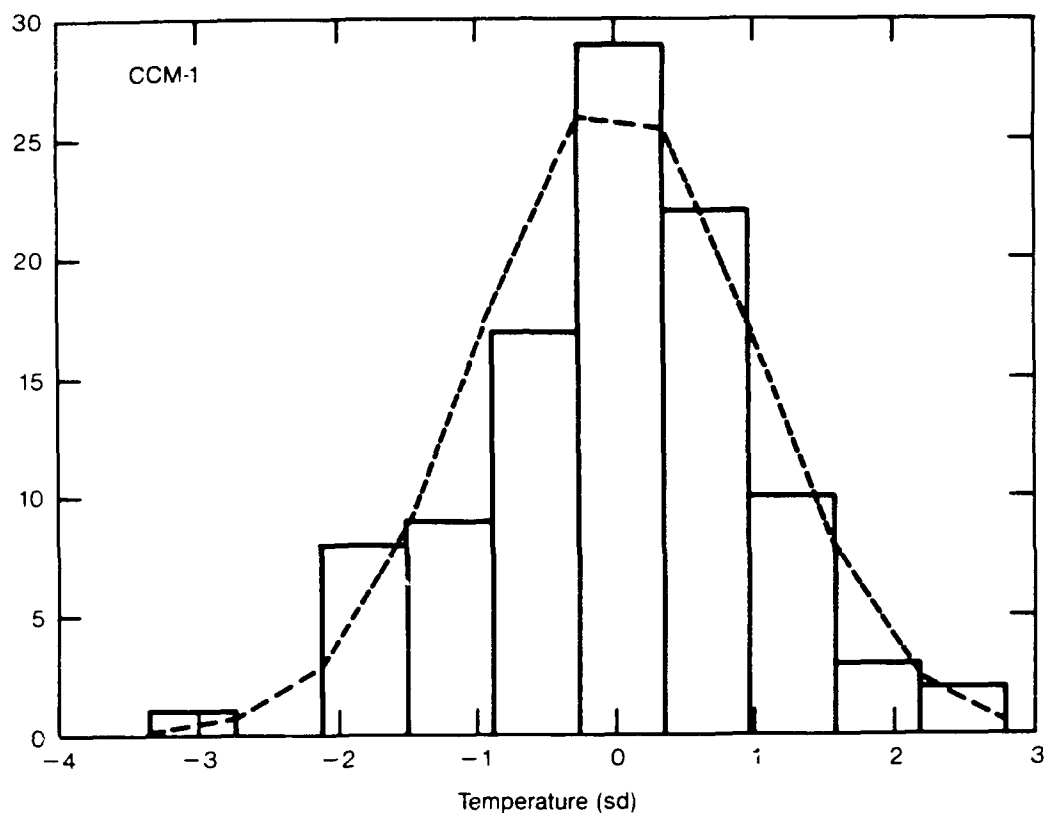
Figure 27. Histogram for the temperature values shown in Figure 26.

Figure 28. Variations of range for the temperature series given in Figure 26. The dashed curve provides the expected value of the range for a normally distributed random variable.

54

(see Figures 29 and 30). The CCM-1 results do not show the red spectrum characteristic of observed temperature records and exhibited by the Lorenz model. In part the difference arises from the treatment of the oceans. The energetics of the ocean determines the "slow" physics. The prescription of the boundary condition over the ocean as the climatological mean ensures that the model reproduces seasonal fluctuations approximately correctly. The CCM-1 representation does not exhibit longer term and higher amplitude fluctuations seen in nature. Since the ultimate goal of models is to "predict" long term changes of climate flowing from anthropogenic compositional changes, the lack of faithful representation of slow changes is a major deficiency.

The 101-year run of CCM-1 shows a persistence index between 0.1 and 0.2 depending on the method used to estimate persistence. These low values and the shortness of the record raise questions as to whether the estimated persistence is statistically significant.

5.3 Concluding Observations

The analysis presented above illustrates the usefulness of having a single number P, the index of persistence, characterize the long term behavior of a time series. The existence of a large positive value of P over a range of time scales indicates that predictability is possible in the sense that runs or trends are likely to be preserved over these time intervals. Long term trends such as that due to changing atmospheric composition give rise to persistence. Sharp lines in the low frequency part of the spectrum also result in persistence. On short time scales, persistence is due to the existence of
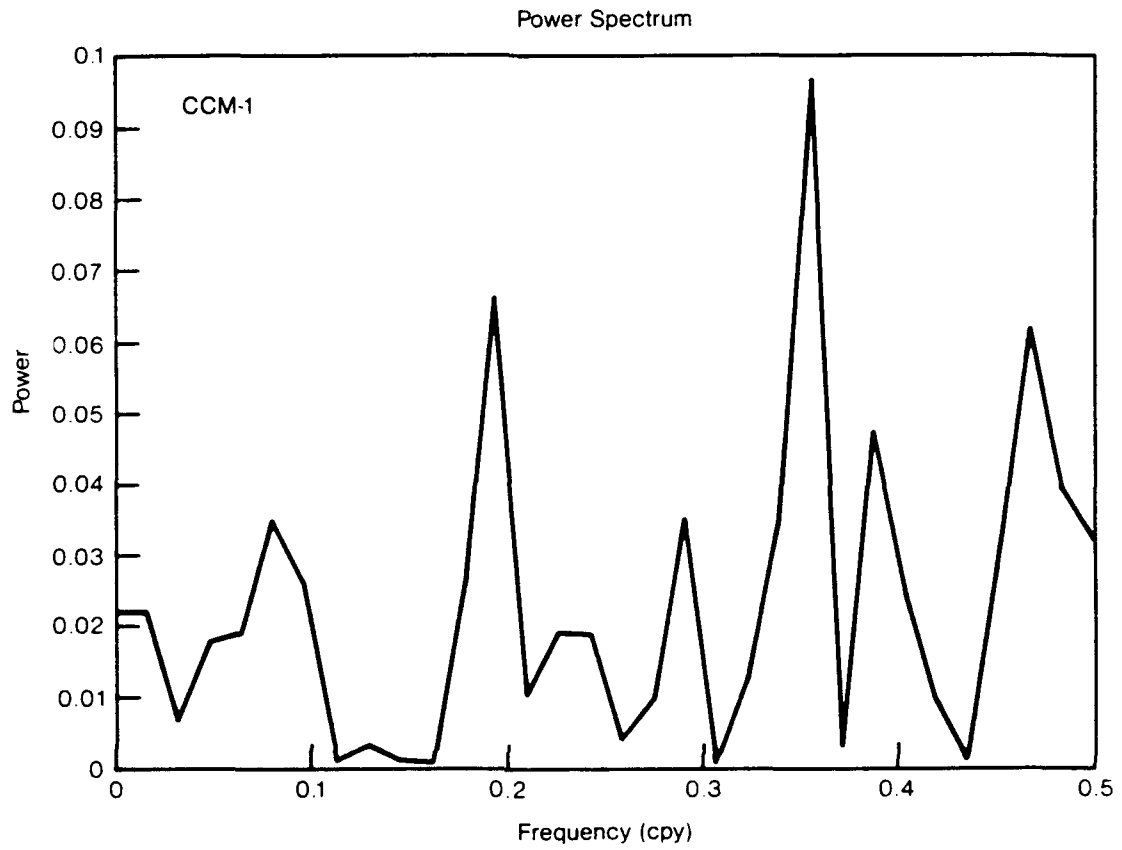
Figure 29. Power spectrum for the global annual average surface temperature shown in Figure 26.

56

Cumulative Spectrum
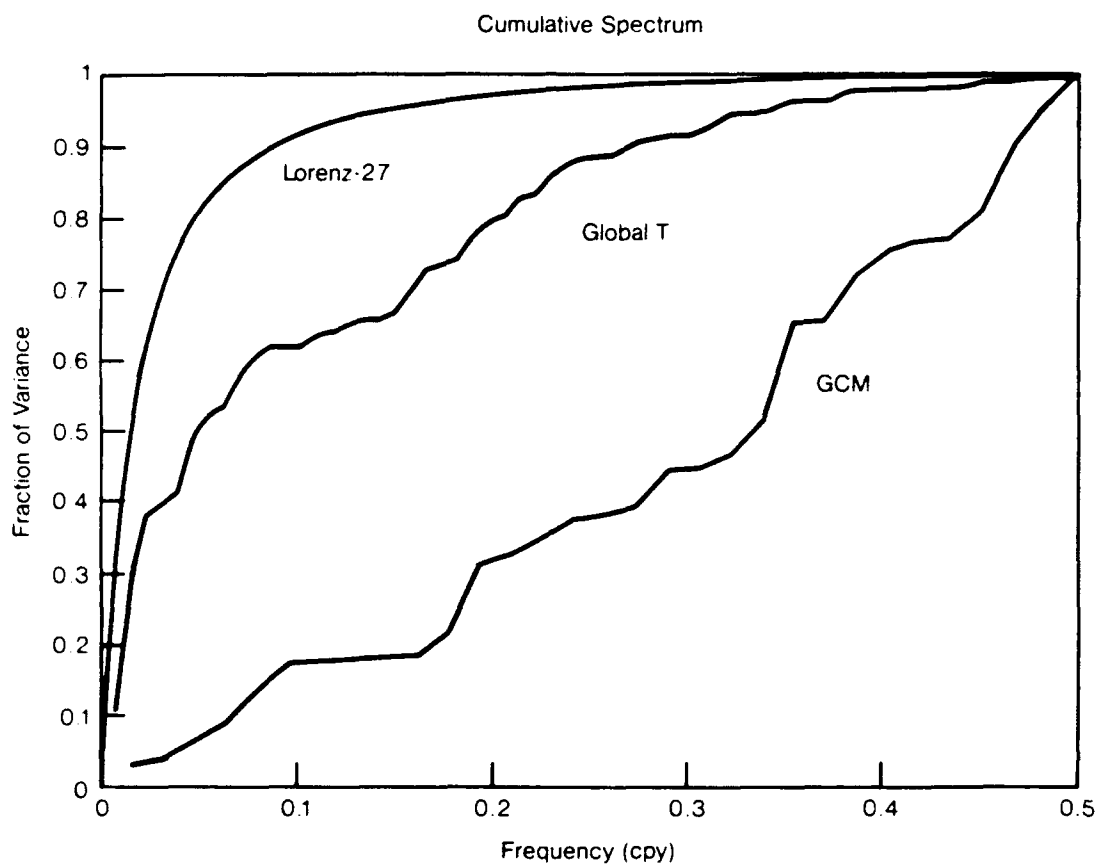
Figure 30. Cumulative spectrum for global annual average temperature obtained for Lorenz-27 model, the observed temperature (see Figure 21) and the GCM results given in Figure 26.

57

large coherent structures in the atmosphere. In the case of climate, large scale coherent structures in the ocean may give rise to persistence over decades or longer.

The analysis of observed global annual average temperature records shows that these records have a strong persistence. While the records are short, only about a century, the persistence is highly statistically significant. The analysis of one GCM record of 101 years does not show persistence or a strong red spectrum and thus differs significantly from the observed global annual average temperature records. The lack of energy at low frequencies is most probably due to a failure to deal with the oceans in an adequate way.

REFERENCES

1. Abarbanel, H., S. Koonin, H. Levine, G. MacDonald, and O. Rothaus (1991b) Issues in predictability, JASON Report JSR-90-320, MITRE Corp., McLean, VA.

2. Abarbanel, H., S. Koonin, H. Levine, G. MacDonald, and O. Rothaus (1991a) Statistics of extremes with application to climate, JASON Report JSR-90-305, MITRE Corp., McLean, VA.

3. David, H. (1981) *Order Statistics*, J. Wiley, New York.

4. Feller, W. (1951) The asymptotic distribution of the range of sums of independent variables. *Ann. Math. Stat.*, **22**, 427-432.

5. Hansen, J., and S. Lebedeff (1988) Global surface temperature; update through 1987. *Geophys. Res. Letters*, **15**, 323-326

6. Huber, P. (1972) Robust statistics: a review, *Ann. Math. Stat.*, **43**, 1041-1067.

7. Hurst, H. (1951) Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civ. Engineers*, **116**, 770-808.

8. Jones, P., S. Raper, R. Bradley, H. Diaz, P. Kelley, and T. Wigley (1986a) Northern hemisphere surface air temperature variations: 1851-1984. *J. Clim. Appl. Meteorol.*, **25**, 161-179.

9. Jones, P., S. Raper, and T. Wigley (1986b) Southern hemisphere surface air temperature variations, *J. Clim. Appl. Meteorol*, **25**, 1213-1230.

10. Jones, P., and T. Wigley (1990) Global warming trends, *Sci. Amer.*, **263**, 84-91.

11. Kolmogorov. A. (1940) Wiener spiral and some other interesting curves in Hilbert space. *Dokl. Acad. Nauk SSSR*, **26**, 115-118.

12. Leadbetter, M., C. Lindgren, and H. Rootzen (1980) *Extreme and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

13. Levine, H., G. MacDonald, O. Rothaus, and F. Zachariasen (1990) Detecting the greenhouse signal, JASON Report JSR-89-330, MITRE Corp., McLean, VA.

14. Levy, P. (1953) Random functions: General theory with special reference to Laplacian random functio.. *Univ. Cal. Publ. Stat.*, **1**, 351-396.

15. Manley, G. (1974) Central England temperatures: Monthly means 1659-1973, *Quart. J. Roy. Meteor. Soc.*, **100**, 389-405.

16. Park, S., and K. Miller (1988) Random number generators: Good ones are hard to find. *Comm ACM*, **32**, 1192-1201.

DISTRIBUTION LIST

Director Ames Laboratory   [2]
Iowa State University
Ames, IA 50011

Mr John M Bachkosky
Deputy DDR&E
The Pentagon
Room 3E114
Washington, DC 20301

Dr Joseph Ball
Central Intelligence Agency
Washington, DC 20505

Dr Arthur E Bisson
DASWD (OASN/RD&A)
The Pentagon
Room 5C675
Washington, DC 20350-1000

Dr Albert Brandenstein
Chief Scientist
Office of Natl Drug Control Policy
Executive Office of the President
Washington, DC 20500

Mr Edward Brown
Assistant Director
Nuclear Monitoring Research Office
DARPA
3701 North Fairfax Drive
Arlington, VA 22203

Dr Herbert L Buchanan III
Director
DARPA/DSO
3701 North Fairfax Drive
Arlington, VA 22203

Dr Curtis G Callan Jr
Physics Department
PO Box 708
Princeton University
Princeton, NJ 08544

Dr Ferdinand N Cirillo Jr
Central Intelligence Agency
Washington, DC 20505

Brig Gen Stephen P Condon
Deputy Assistant Secretary
Management Policy &
Program Integration
The Pentagon Room 4E969
Washington, DC 20330-1000

Ambassador Henry F Cooper
Director/SDIO-D
Room 1E1081
The Pentagon
Washington, DC 20301-7100

DARPA
RMO/Library
3701 North Fairfax Drive
Arlington, VA 22209-2308

61

DISTRIBUTION LIST

Mr John Darrah
Senior Scientist and Technical Advisor
HQAF SPACOM/CN
Peterson AFB, CO 80914-5001

Col Doc Dougherty
DARPA/DIRO
3701 North Fairfax Drive
Arlington, VA 22203

DTIC [2]
Defense Technical Information Center
Cameron Station
Alexandria, VA 22314

Mr John N Entzminger
Chief, Advance Technology
DARPA/ASTO
3701 North Fairfax Drive
Arlington, VA 22203

CAPT Kirk Evans
Director Undersea Warfare
Space & Naval Warfare Sys Cmd
Code PD-80
Department of the Navy
Washington, DC 20363-5100

Mr F Don Freeburn
US Department of Energy
Code ER-33
Mail Stop G-236
Washington, DC 20585

Dr Dave Galas
Associate Director for
Health & Environmental Research
ER-70/GTN
US Department of Energy
Washington, DC 20585

Dr S William Gouse
Sr Vice President and General Manager
The MITRE Corporation
Mail Stop Z605
7525 Colshire Drive
McLean, VA 22102

LTGEN Robert D Hammond
CMDR & Program Executive Officer
US Army/CSSD-ZA
Strategic Defense Command
PO Box 15280
Arlington, VA 22215-0150

Mr Thomas H Handel
Office of Naval Intelligence
The Pentagon
Room 5D660
Washington, DC 20350-2000

Maj Gen Donald G Hard
Director of Space and SDI Programs
Code SAF/AQS
The Pentagon
Washington, DC 20330-1000

Dr Robert G Henderson
Director
JASON Program Office
The MITRE Corporation
7525 Colshire Drive Z561
McLean, VA 22102

Dr Barry Horowitz
President and Chief Executive Officer
The MITRE Corporation
Burlington Road
Bedford, MA 01730

Dr William E Howard III [2]
Director For Space
and Strategic Technology
Office/Assistant Secretary of the Army
The Pentagon Room 3E474
Washington, DC 20310-0103

Dr Gerald J Iafrate
US Army Research Office
PO Box 12211
4300 South Miami Boulevard
Research Triangle Park, NC 27709-2211

Technical Information Center [2]
US Department of Energy
PO Box 62
Oak Ridge, TN 37830

JASON Library [5]
The MITRE Corporation
Mail Stop W002
7525 Colshire Drive
McLean, VA 22102

Dr George Jordy [25]
Director for Program Analysis
US Department of Energy
MS ER30 Germantown
OER
Washington, DC 20585

Dr O'Dean P Judd
Los Alamos National Lab
Mail Stop A-110
Los Alamos, NM 87545

Technical Librarian [2]
Argonne National Laboratory
9700 South Cass Avenue
Chicago, IL 60439

Research Librarian [2]
Brookhaven National Laboratory
Upton, NY 11973

Technical Librarian [2]
Los Alamos National Laboratory
PO Box 1663
Los Alamos, NM 87545

Technical Librarian [2]
Pacific Northwest Laboratory
PO Box 999
Battelle Boulevard
Richland, WA 99352

Technical Librarian [2]
Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185

DISTRIBUTION LIST

Technical Librarian   [2]
Sandia National Laboratories
PO Box 969
Livermore, CA 94550

Technical Librarian   [2]
Lawrence Berkeley Laboratory
One Cyclotron Road
Berkeley, CA 94720

Technical Librarian   [2]
Lawrence Livermore Nat'l Lab
PO Box 808
Livermore, CA 94550

Technical Librarian   [2]
Oak Ridge National Laboratory
Box X
Oak Ridge, TN 37831

Chief Library Branch   [2]
AD-234.2 FORS
US Department of Energy
Washington, DC 20585

Dr Gordon J MacDonald
Institute on Global Conflict
& Cooperation
UCSD/0518
9500 Gilman Drive
La Jolla, CA 92093-0518

Mr Robert Madden   [2]
Department of Defense
National Security Agency
Attn R-9 (Mr Madden)
Ft George G Meade, MD 20755-6000

Dr Arthur F Manfredi Jr   [10]
OSWR
Central Intelligence Agency
Washington, DC 20505

Mr Joe Martin
Director
OUSD(A)/TWP/NW&M
Room 3D1048
The Pentagon
Washington, DC 20301

Mr Ronald Murphy
DARPA/ASTO
3701 North Fairfax Drive
Arlington, VA 22203-1714

Dr Julian C Nall
Institute for Defense Analyses
1801 North Beauregard Street
Alexandria, VA 22311

Dr Gordon C Oehler
Central Intelligence Agency
Washington, DC 20505

DISTRIBUTION LIST

Oak Ridge Operations Office
Procurement and Contracts Division
US Department of Energy
(DOE IA No DE-AI05-90ER30174)
PO Box 2001
Oak Ridge, TN 37831-8757

Dr Peter G Pappas
Chief Scientist
US Army Strategic Defense Command
PO Box 15280
Arlington, VA 22215-0260

Dr Aristedes Patrinos [20]
Director of Atmospheric
& Climate Research
ER-74/GTN
US Department of Energy
Washington, DC 20585

Dr Bruce Pierce
USD(A)/D S
Room 3D136
The Pentagon
Washington, DC 20301-3090

Mr John Rausch [2]
Division Head 06 Department
NAVOPINTCEN
4301 Suitland Road
Washington, DC 20390

Records Resources
The MITRE Corporation
Mailstop W115
7525 Colshire Drive
McLean, VA 22102

Dr Fred E Saalfeld
Director
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

Dr John Schuster
Technical Director of Submarine
and SSBN Security Program
Department of the Navy OP-02T
The Pentagon Room 4D534
Washington, DC 20350-2000

Dr Barbara Seiders
Chief of Research
Office of Chief Science Advisor
Arms Control & Disarmament Agency
320 21st Street NW
Washington, DC 20451

Dr Philip A Selwyn [2]
Director
Office of Naval Technology
Room 907
800 North Quincy Street
Arlington, VA 22217-5000

Superintendent
CODE 1424
Attn Documents Librarian
Naval Postgraduate School
Monterey, CA 93943

Dr George W Ullrich [3]
Deputy Director
Defense Nuclear Agency
6801 Telegraph Road
Alexandria, VA 22310

DISTRIBUTION LIST

Ms Michelle Van Cleave
Asst Dir/National Security Affairs
Office/Science and Technology Policy
New Executive Office Building
17th and Pennsylvania Avenue
Washington, DC 20506

Mr Richard Vitali
Director of Corporate Laboratory
US Army Laboratory Command
2800 Powder Mill Road
Adelphi, MD 20783-1145

Dr Edward C Whitman
Dep Assistant Secretary of the Navy
C3I Electronic Warfare & Space
Department of the Navy
The Pentagon 4D745
Washington, DC 20350-5000

Mr Donald J. Yockey
U/Secretary of Defense
For Acquisition
The Pentagon Room 3E933
Washington, DC 20301-3000

Dr Linda Zall
Central Intelligence Agency
Washington, DC 20505

Mr Charles A Zraket
Trustee
The MITRE Corporation
Mail Stop A130
Burlington Road
Bedford, MA 01730